

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/12, C07K 14/475, 16/22, G01N 33/68, C12Q 1/68, A61K 38/18	A1	(11) International Publication Number: WO 98/44112 (43) International Publication Date: 8 October 1998 (08.10.98)
(21) International Application Number: PCT/US98/06022 (22) International Filing Date: 27 March 1998 (27.03.98) (30) Priority Data: 60/042,855 28 March 1997 (28.03.97) US (71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). (74) Agents: STEFFE, Eric, K.; Sterne, Kessler, Goldstein & Fox P.L.L.C., Suite 600, 1100 New York Avenue, N.W., Washington, DC 20005-3934 (US) et al.	(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> <i>With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.</i>	
(54) Title: HUMAN MUSCLE DERIVED GROWTH FACTOR - CARDIAC AND PANCREATIC PROTEIN (CAPP) AND GENE (57) Abstract <p>The present invention relates to a novel Cardiac And Pancreatic Protein (CAPP) which is a member of the muscle derived growth factor superfamily. In particular, isolated nucleic acid molecules are provided encoding the human CAPP protein. CAPP polypeptides are also provided as are vectors, host cells and recombinant methods for producing the same. The invention further relates to screening methods for identifying agonists and antagonists of CAPP activity. Also provided are diagnostic methods for detecting the presence of activated T-cells, and mature heart, pancreas and placental tissues and cells. The CAPP polypeptides can also be employed in cell culture media for regulating cell differentiation and maintaining heart, placenta and pancreas cells.</p>		

DOCID: <wo_9844112A1_1>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroun	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

ISDOCIO: dno_9844112A1.J

WO 98/44112

PCT/US98/06022

HUMAN MUSCLE DERIVED GROWTH FACTOR - CARDIAC AND PANCREATIC PROTEIN (CAPP) AND GENE

*Background of the Invention**Field of the Invention*

5 The present invention relates to a novel muscle derived growth factor. More specifically, isolated nucleic acid molecules are provided encoding human Cardiac And Pancreatic Protein (CAPP). CAPP polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are diagnostic methods for detecting CAPP gene expression and methods for stimulating and inhibiting the growth of certain cells.

10 *Related Art*

Control of cell division is a basic aspect of multicellular existence that depends upon a programmed series of events. One factor in cellular proliferation and its control that has been identified is the presence of various polypeptide growth factors. Growth factors are essential components of growth media for *in vitro* cell culture and are involved in cell survival *in vivo*. Some of the growth factors that have been identified to date include PDGF (platelet-derived growth factor) implicated in the repair of the vascular system *in vivo*; EGF (epidermal growth factor) which acts as a mitogen for cells of ectodermal and mesodermal origin; TGF- α (transforming growth factor) which acts as a mitogen similarly to EGF but can make normal cells grow in agar; TGF- β (transforming growth factor) which is a mitogen for some cells and a growth inhibitor for others; and NGF (nerve growth factor) involved in the development and maintenance of sympathetic and embryonic neurons. Watson *et al.*, Molecular Biology of the Gene, p. 975 (Benjamin/Cummings 1987).

25 It is clear that particular cell types require particular growth factors. Peptide growth factors are produced and secreted from a variety of tissues. The target cells are typically located close to the site of release of the growth factor

WO 98/44112

PCT/US98/06022

2

(paracrine response). In addition to their growth promoting and differentiation-inducing activities, growth factors elicit a wide variety of effects in their target cells and are involved in processes such as inflammation, immune reactions and wound repair. See, Pimentel, E. *Handbook of Growth Factors, Volume 1: General Basics* (CRC Press 1994).

Myocardial hypertrophy refers to a focal or general enlargement of the heart. Normal hypertrophy is a compensatory function to maintain the pumping function of the heart. Abnormal hypertrophy occurs in hypertension, myocardial infarction, valve disease and cardiomyopathy. Simpson, P.C. *Heart Failure* 5:113 (1989). Cardiac myocytes have been shown to be targets for the effects of peptide growth factors on differentiated gene expression. Stimulation of the α_1 -adrenergic receptor induces hypertrophy of cultured cardiac myocytes and produces specific changes in gene expression at the level of transcription. Simpson, P.C. "Cardiac Myocyte Hypertrophy," *Molecular Biology of the Cardiovascular System*, Roberts, R. et al. ed.:125-133 (1990). In cardiac myocytes, the growth factors TGF- β 1 and basic FGF concomitantly elicit complex and heterogeneous responses: selective inhibition of certain adult transcripts, concurrent with the up-regulation of "fetal" contractile protein genes. Schneider et al., "Oncogenes and Myogenesis," *Molecular Biology of the Cardiovascular System*, Roberts, R. et al. ed.:63-71 (1990).

Monitoring of growth factor gene expression in myocytes and the other cells of the heart, including connective tissue, would be useful in detecting and studying abnormal hypertrophy both *in vitro* and *in vivo*. Organ and clonal cell systems have been developed to analyze cardiomyogenic differentiation. See, for example, Bader, D. et al., *Molecular Biology of the Cardiovascular System*, Roberts, R. et al. ed.:41-49 (1990). Differentiation in these systems can be monitored by *in vitro* analysis of cardiac myogenesis and monoclonal antibodies that have been raised against muscle-specific proteins.

Additionally, polypeptide growth factors are very important cell culture reagents for stimulating cellular growth and aiding survival of the cells *in vitro*.

The search continues to exist for polypeptides that stimulate and/or inhibit

WO 98/44112

PCT/US98/06022

3

growth of particular cells for both *in vitro* and *in vivo* uses. In addition, the search continues for novel tissue specific markers that can be employed qualitatively to help identify a particular cell or tissue type and employed qualitatively to assess whether cells, tissues or organs are abnormal in their expression of a particular polypeptide.

Summary of the Invention

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding CAPP polypeptide having the amino acid sequence is shown in Figure 1 (SEQ ID NO:2) or the amino acid sequence encoded by the cDNA clone deposited in a bacterial host which was deposited on September 23, 1996 at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, and given accession number 97729. The nucleotide sequence determined by sequencing the deposited CAPP clone, which is shown in Figure 1 (SEQ ID NO:1), contains an open reading frame encoding a polypeptide of 397 amino acid residues, including an initiation codon at positions 1-3, with a leader sequence of about 32 amino acid residues, and a predicted molecular weight of about 40 kDa. The amino acid sequence of the mature CAPP protein is shown in Figure 1, amino acid residues 1-365 in SEQ ID NO:2.

Thus, one aspect of the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding the CAPP polypeptide having the complete amino acid sequence in Figure 1 (SEQ ID NO:2); (b) a nucleotide sequence encoding the mature CAPP polypeptide having the amino acid sequence at positions 33-397 in Figure 1 or 1-365 in SEQ ID NO:2; (c) a nucleotide sequence encoding the CAPP polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; (d) a nucleotide sequence encoding the mature CAPP polypeptide having the amino acid sequence encoded by the cDNA clone

DOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

4

contained in ATCC Deposit No. 97729; and (e) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c) or (d) above.

Further embodiments of the invention include isolated nucleic acid molecules that comprise a polynucleotide having a nucleotide sequence at least
5 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical, to any of the nucleotide sequences in (a), (b), (c), (d) or (e), above, or a polynucleotide which hybridizes under stringent hybridization conditions to a polynucleotide in (a), (b), (c), (d) or (e), above. This polynucleotide which hybridizes does not hybridize under stringent hybridization conditions to a
10 polynucleotide having a nucleotide sequence consisting of only A residues or of only T residues. An additional nucleic acid embodiment of the invention relates to an isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of a CAPP polypeptide having an amino acid sequence in (a), (b), (c) or (d), above.

15 The present invention also relates to recombinant vectors, which include the isolated nucleic acid molecules of the present invention, and to host cells containing the recombinant vectors, as well as to methods of making such vectors and host cells and for using them for production of CAPP polypeptides or peptides by recombinant techniques.

20 The invention further provides an isolated CAPP polypeptide having an amino acid sequence selected from the group consisting of: (a) the amino acid sequence of the CAPP polypeptide having the complete 397 amino acid sequence, including the leader sequence shown in Figure 1 (SEQ ID NO:2); (b) the amino acid sequence of (b) the amino acid sequence of the mature CAPP polypeptide
25 (without the leader) having the amino acid sequence at positions 1-365 in SEQ ID NO:2; (c) the amino acid sequence of the CAPP polypeptide having the complete amino acid sequence, including the leader, encoded by the cDNA clone contained in ATCC Deposit No. 97729; and (d) the amino acid sequence of the mature CAPP polypeptide having the amino acid sequence encoded by the cDNA
30 clone contained in ATCC Deposit No. 97729. The polypeptides of the present invention also include polypeptides having an amino acid sequence with at least

ISDOCD: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

5

90% similarity, and more preferably at least 95% similarity to those described in (a), (b), (c) or (d) above, as well as polypeptides having an amino acid sequence at least 80% identical, more preferably at least 90% identical, and still more preferably 95%, 96%, 97%, 98% or 99% identical to those above.

5 An additional embodiment of this aspect of the invention relates to a peptide or polypeptide which has the amino acid sequence of an epitope-bearing portion of a CAPP polypeptide having an amino acid sequence described in (a), (b), (c) or (d), above. Peptides or polypeptides having the amino acid sequence of an epitope-bearing portion of a CAPP polypeptide of the invention include
10 portions of such polypeptides with at least six or seven, preferably at least nine, and more preferably at least about 30 amino acids to about 50 amino acids, although epitope-bearing polypeptides of any length up to and including the entire amino acid sequence of a polypeptide of the invention described above also are included in the invention. In another embodiment, the invention provides an
15 isolated antibody that binds specifically to a CAPP polypeptide having an amino acid sequence described in (a), (b), (c) or (d) above.

The invention further provides methods for isolating antibodies that bind specifically to a CAPP polypeptide having an amino acid sequence as described herein. Such antibodies are useful diagnostically or therapeutically as describe
20 below.

The expression of CAPP protein is expected to be necessary for the survival and maintenance of mature muscle cells, especially heart, placenta and pancreas tissue. Under certain conditions the CAPP protein is expected to act with other growth factors to modulate, e.g. block, the proliferation of mature heart
25 and pancreas cells. Under certain conditions the CAPP protein is expected to act with other growth factors to program the differentiation of immature cells into cardiac or pancreatic cells. These functional properties of this peptide can be exploited *in vivo* in a number of useful ways. An antagonist of the CAPP protein may be useful in allowing mature muscle cells, such as myocytes to replicate and
30 divide, something that does not occur in most normal myocytes. *In vitro* the CAPP protein can be employed to cause embryonic cells to differentiate into

DOCID: <WO_9844112A1_J>

WO 98/44112

PCT/US98/06022

6

cardiac cells and to maintain cell cultures of cardiac, pancreatic or placental cells.

The CAPP gene was discovered in an activated T-cell cDNA library. CAPP gene expression and translation can be used as a marker to detect activated T-cells. Monitoring T cells activation is useful for a number of *in vitro* diagnostic purposes, including studying the effects of candidate drugs on the immune system, and determining whether the T cells of a subject have been activated by analyzing a blood sample taken from the subject or by assessing activity in an *in vitro* screening test.

The present inventors have discovered that CAPP is highly expressed in adult heart, pancreas and placenta tissue. For a number of disorders of smooth muscle tissue in the heart, pancreas or placenta, it is believed that significantly higher or lower levels of CAPP gene expression can be detected in certain tissues (e.g., heart, pancreas and placenta) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) taken from an individual having such a disorder, relative to a "standard" CAPP gene expression level, i.e., the CAPP expression level in tissue or bodily fluids from an individual not having a disorder of the heart, pancreas or placenta. Thus, the invention provides a diagnostic method useful during diagnosis of an internal organ disorder, wherein said disorder relates to the smooth muscle tissue of the heart, pancreas or placenta, which involves: (a) assaying CAPP gene expression level in cells or body fluid of an individual; (b) comparing the CAPP gene expression level with a standard CAPP gene expression level, whereby an increase or decrease in the assayed CAPP gene expression level compared to the standard expression level is indicative of one of said disorders.

Additionally, this CAPP gene expression can be employed as a marker to determine the presence of mature, terminally differentiated organ tissue, especially heart, pancreatic and placental tissue. Such a marker possesses practical utility in monitoring the growth of heart, pancreas and placental cells and tissues *ex vivo*. The effects of small molecule drugs and polypeptide growth factors on the development of these cells and tissues can be assessed by monitoring the level of expression of the CAPP gene.

WO 98/44112

PCT/US98/06022

7

5 The present invention also provides a screening method for identifying compounds capable of enhancing or inhibiting a cellular response induced by the CAPP polypeptide, which involves contacting cells which express the CAPP polypeptide with the candidate compound, assaying a cellular response, and comparing the cellular response to a standard cellular response, the standard being assayed when contact is made in absence of the candidate compound; whereby, an increased cellular response over the standard indicates that the compound is an agonist and a decreased cellular response over the standard indicates that the compound is an antagonist.

10 In another aspect, a screening assay for agonists and antagonists is provided which involves determining the effect a candidate compound has on CAPP polypeptide modulation of cellular growth and differentiation. In particular, the method involves contacting a cell culture with CAPP polypeptide and a candidate compound and determining whether CAPP polypeptide increases or decreases cellular differentiation or proliferation in the presence of the candidate compound.

15 An additional aspect of the invention is related to a method for treating an individual in need of an increased level of CAPP activity in the body comprising administering to such an individual a composition comprising a therapeutically effective amount of an isolated CAPP polypeptide of the invention or an agonist thereof.

20 A still further aspect of the invention is related to a method for treating an individual in need of a decreased level of CAPP activity in the body comprising, administering to such an individual a composition comprising a therapeutically effective amount of a CAPP antagonist. Preferred antagonists for use in the present invention are CAPP-specific antibodies.

Brief Description of the Figures

30 Figure 1 shows the nucleotide (SEQ ID NO:1) and deduced amino acid (SEQ ID NO:2) sequences of CAPP. The protein has a leader sequence of about

DOCID: WO_9844112A1_1

WO 98/44112

PCT/US98/06022

8

32 amino acid residues (underlined) and a deduced molecular weight of about 40 kDa. The predicted amino acid sequence of the mature CAPP protein is also shown in Figure 1 (SEQ ID NO:2).

5 Figure 2 shows the regions of similarity between the amino acid sequences of the CAPP protein and *Drosophila* "brainiac" gene (SEQ ID NO:3).

Figure 3 shows an analysis of the CAPP amino acid sequence. Alpha, beta, turn and coil regions; hydrophilicity and hydrophobicity; amphipathic regions; flexible regions; antigenic index and surface probability are shown. In the "Antigenic Index - Jameson-Wolf" graph, the following amino acid residues in Figure 1 correspond to the shown highly antigenic regions of the CAPP protein.

10 Figure 4 shows a schematic representation of the pHE4a expression vector (SEQ ID NO:8). The locations of the kanamycin resistance marker gene, the multiple cloning site linker region, the oriC sequence, and the *lacIq* coding sequence are indicated.

15 Figure 5 shows the nucleotide sequence of the regulatory elements of the pHE4a promoter (SEQ ID NO:9). The two *lac* operator sequences, the Shine-Delgarno sequence (S/D), and the terminal *HindIII* and *NdeI* restriction sites (italicized) are indicated.

20 Detailed Description

The present invention provides isolated nucleic acid molecules comprising a polynucleotide encoding a CAPP polypeptide having the amino acid sequence shown in Figure 1 (SEQ ID NO:2), which was determined by sequencing a cloned cDNA. The CAPP protein of the present invention shares sequence homology with *Drosophila* "brainiac" gene (Figure 2) (SEQ ID NO:3). The nucleotide sequence shown in Figure 1 (SEQ ID NO:1) was obtained by sequencing the HTAAW41 clone, which was deposited on September 23, 1996 at the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, and given accession number 97729. The deposited clone is contained in the pBluescript SK(-) plasmid (Stratagene, La Jolla, CA).

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

9

Nucleic Acid Molecules

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer (such as the Model 373 from Applied Biosystems, Inc.), and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

Unless otherwise indicated, each "nucleotide sequence" set forth herein is presented as a sequence of deoxyribonucleotides (abbreviated A, G, C and T). However, by "nucleotide sequence" of a nucleic acid molecule or polynucleotide is intended, for a DNA molecule or polynucleotide, a sequence of deoxyribonucleotides, and for an RNA molecule or polynucleotide, the corresponding sequence of ribonucleotides (A, G, C and U), where each thymidine deoxyribonucleotide (T) in the specified deoxyribonucleotide sequence is replaced by the ribonucleotide uridine (U). For instance, reference to an RNA molecule having the sequence of SEQ ID NO:1 set forth using deoxyribonucleotide abbreviations is intended to indicate an RNA molecule having a sequence in which each deoxyribonucleotide A, G or C of SEQ ID NO:1

:DOCID: <WO_984112A1.J>

WO 98/44112

PCT/US98/06022

10

has been replaced by the corresponding ribonucleotide A, G or C, and each deoxyribonucleotide T has been replaced by a ribonucleotide U.

Using the information provided herein, such as the nucleotide sequence in Figure 1, a nucleic acid molecule of the present invention encoding a CAPP polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA as starting material. Illustrative of the invention, the nucleic acid molecule described in Figure 1 (SEQ ID NO:1) was discovered in a cDNA library derived from activated T cells. By Northern blot analysis it has been determined that this gene is abundant in adult heart and pancreas, with low amounts in placenta, lung, liver, skeletal muscle, kidney, spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocytes. The gene was identified by database distribution in activated T cells (3), CD34 positive cells, Ntera2 cells 14 days after RA stimulation, kidney cortex, adult heart, Jurkat cells and small intestine. The determined nucleotide sequence of the CAPP cDNA of Figure 1 (SEQ ID NO:1) contains an open reading frame encoding a protein of 397 amino acid residues, with an initiation codon at positions 233-236 of the nucleotide sequence in Figure 1 (SEQ ID NO:1), a predicted leader sequence of about 32 amino acid residues, and a deduced molecular weight of about 40 kDa. The amino acid sequence of the predicted mature CAPP is shown in 1 to residue 365 in SEQ ID NO:2. The CAPP protein shown in Figure 1 (SEQ ID NO:2) is about 27.8% identical and about 48.4% similar to *Drosophila* "brainiac" protein (Figure 2). This protein interacts with the EGF receptor pathway in follicle cell development (Goode *et al.*, *Development* 116:177-192 (1992)).

As one of ordinary skill would appreciate, due to the possibilities of sequencing errors discussed above, as well as the variability of cleavage sites for leaders in different known proteins, the actual CAPP polypeptide encoded by the deposited cDNA comprises about 397 amino acids, but may be anywhere in the range of 385-410 amino acids; and the actual leader sequence of this protein is about 32 amino acids, but may be anywhere in the range of about 25 to about 40 amino acids.

WO 98/44112

PCT/US98/06022

11

As indicated, nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced synthetically. The DNA may be double-stranded or single-stranded. Single-stranded DNA or RNA
5 may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

By "isolated" nucleic acid molecule(s) is intended a nucleic acid molecule, DNA or RNA, which has been removed from its native environment. For example, recombinant DNA molecules contained in a vector are considered
10 isolated for the purposes of the present invention. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the DNA molecules of the present invention. Isolated nucleic acid molecules according to
15 the present invention further include such molecules produced synthetically.

Isolated nucleic acid molecules of the present invention include DNA molecules comprising an open reading frame (ORF) with an initiation codon at positions 233-236 of the nucleotide sequence shown in Figure 1 (SEQ ID NO:1); DNA molecules comprising the coding sequence for the mature CAPP protein
20 shown in Figure 1 (last 365 amino acids) (SEQ ID NO:2); and DNA molecules which comprise a sequence substantially different from those described above but which, due to the degeneracy of the genetic code, still encode the CAPP protein. Of course, the genetic code is well known in the art. Thus, it would be routine for one skilled in the art to generate the degenerate variants described above.

25 In addition, the invention present inventors have identified the following cDNA clones related to extensive portions of SEQ ID NO:1: HAHAA70F (SEQ ID NO:10), HTAAW41R (SEQ ID NO:11), HTABE60R (SEQ ID NO:12), HJUBA94R (SEQ ID NO:13), HSIBA68R (SEQ ID NO:14), and HSIBA68F (SEQ ID NO:15).

30 In addition, the invention present inventors have identified the following public cDNA clones related to extensive portions of SEQ ID NO:1: AA773646

WO 98/44112

PCT/US98/06022

12

(SEQ ID NO:16), AA449869 (SEQ ID NO:17), N66915 (SEQ ID NO:18),
H93550 (SEQ ID NO:19), W26453 (SEQ ID NO:20), H13125 (SEQ ID NO:21),
N51037 (SEQ ID NO:22), N58174 (SEQ ID NO:23), R74552 (SEQ ID NO:24),
R82733 (SEQ ID NO:25), H78875 (SEQ ID NO:25), H47991 (SEQ ID NO:26),
5 R74454 (SEQ ID NO:27), C20629 (SEQ ID NO:28), AA310578 (SEQ ID
NO:29), AA263148 (SEQ ID NO:30), H00589 (SEQ ID NO:31), R31680 (SEQ
ID NO:32), AA381631 (SEQ ID NO:33), H80116 (SEQ ID NO:34), H47990
(SEQ ID NO:35), AA381830 (SEQ ID NO:36), AA377082 (SEQ ID NO:37),
H13126 (SEQ ID NO:38), R31722 (SEQ ID NO:39), AA377081 (SEQ ID
10 NO:40), and D87736 (SEQ ID NO:41).

In another aspect, the invention provides isolated nucleic acid molecules
encoding the CAPP polypeptide having an amino acid sequence encoded by the
cDNA clone contained in the plasmid deposited as ATCC Deposit No. 97729 on
September 23, 1996. Preferably, this nucleic acid molecule will encode the
15 mature polypeptide encoded by the above-described deposited cDNA clone. The
invention further provides an isolated nucleic acid molecule having the nucleotide
sequence shown in Figure 1 (SEQ ID NO:1) or the nucleotide sequence of the
CAPP cDNA contained in the above-described deposited clone, or a nucleic acid
molecule having a sequence complementary to one of the above sequences. Such
20 isolated molecules, particularly DNA molecules, are useful as probes for gene
mapping, by *in situ* hybridization with chromosomes, and for detecting
expression of the CAPP gene in human tissue, for instance, by Northern blot
analysis.

The present invention is further directed to fragments of the isolated
25 nucleic acid molecules described herein. By a fragment of an isolated nucleic
acid molecule having the nucleotide sequence of the deposited cDNA or the
nucleotide sequence shown in Figure 1 (SEQ ID NO:1) is intended fragments at
least about 15 nt, and more preferably at least about 20 nt, still more preferably
at least about 30 nt, and even more preferably, at least about 40 nt in length which

SEQUENCE: <WO_9844112A1_1_>

WO 98/44112

PCT/US98/06022

13

are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000, 1025, 1050, 1075, 1100, 1125, 1150, 1175, 1200, 1225, 1250, 1275, 1300, 1325, 1350, 1375, 1400, 1425, 1450, 1475, 1500, 1525, 1550, 1575, 1600, 1625, 1650, 1675, 1700, 1725, 1750, 1775, 1800, 1825, 1850, 1875, 1900, 1925, 1950, 1975, 2000, 2025, 2050, 2075, 2100, 2125, 2150, 2175, 2200, 2225, 2250, 2275, 2300, 2325, 2350, 2375, 2400, 2425, 2450, 2475, 2500, 2525, 2550, 2575, 2600, 2625, 2650, 2675, 2700, 2725 and 2740 nt in length are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the deposited cDNA or as shown in Figure 1 (SEQ ID NO:1). By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequence of the deposited cDNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1). Since the gene has been deposited and the nucleotide sequence shown in Figure 1 (SEQ ID NO:1) is provided, generating such DNA fragments would be routine to the skilled artisan. For example, restriction endonuclease cleavage or shearing by sonication could easily be used to generate fragments of various sizes. Alternatively, such fragments could be generated synthetically.

Preferred nucleic acid fragments of the present invention include nucleic acid molecules encoding epitope-bearing portions of the CAPP protein. In particular, such nucleic acid fragments of the present invention include nucleic acid molecules encoding: a polypeptide comprising amino acid residues from about -32 to about -22 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about -4 to about 40 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 46 to about 57 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 62 to about 73 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 78 to about 87 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 92 to about 110 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about

300000: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

14

119 to about 144 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 152 to about 186 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 200 to about 219 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 230 to about 240 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 248 to about 258 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 314 to about 336 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 344 to about 353 in SEQ ID NO:2. The inventors have determined that the above polypeptide fragments are antigenic regions of the CAPP protein. Methods for determining other such epitope-bearing portions of the CAPP protein are described in detail below.

In another aspect, the invention provides an isolated nucleic acid molecule comprising a polynucleotide which hybridizes under stringent hybridization conditions to a portion of the polynucleotide in a nucleic acid molecule of the invention described above, for instance, the cDNA clone contained in ATCC Deposit 97729. By "stringent hybridization conditions" is intended overnight incubation at 42°C in a solution comprising: 50% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1 x SSC at about 65°C.

By a polynucleotide which hybridizes to a "portion" of a polynucleotide is intended a polynucleotide (either DNA or RNA) hybridizing to at least about 15 nucleotides (nt), and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably about 30-70 nt of the reference polynucleotide. These are useful as diagnostic probes and primers as discussed above and in more detail below.

Of course, polynucleotides hybridizing to a larger portion of the reference polynucleotide (e.g., the deposited cDNA clone), for instance, a portion 50-750 nt in length, or even to the entire length of the reference polynucleotide, are also useful as probes according to the present invention, as are polynucleotides corresponding to most, if not all, of the nucleotide sequence of the deposited

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

15

cDNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1). By a portion of a polynucleotide of "at least 20 nt in length," for example, is intended 20 or more contiguous nucleotides from the nucleotide sequence of the reference polynucleotide (e.g., the deposited cDNA or the nucleotide sequence as shown in Figure 1 (SEQ ID NO:1)). As indicated, such portions are useful diagnostically either as a probe according to conventional DNA hybridization techniques or as primers for amplification of a target sequence by the polymerase chain reaction (PCR), as described, for instance, in *Molecular Cloning, A Laboratory Manual*, 2nd. edition, edited by Sambrook, J., Fritsch, E. F. and Maniatis, T., (1989), Cold Spring Harbor Laboratory Press, the entire disclosure of which is hereby incorporated herein by reference.

Since a CAPP cDNA clone has been deposited and its determined nucleotide sequence is provided in Figure 1 (SEQ ID NO:1), generating polynucleotides which hybridize to a portion of the CAPP cDNA molecule would be routine to the skilled artisan. For example, restriction endonuclease cleavage or shearing by sonication of the CAPP cDNA clone could easily be used to generate DNA portions of various sizes which are polynucleotides that hybridize to a portion of the CAPP cDNA molecule. Alternatively, the hybridizing polynucleotides of the present invention could be generated synthetically according to known techniques. Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of the CAPP cDNA shown in Figure 1 (SEQ ID NO:1)), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

As indicated, nucleic acid molecules of the present invention which encode a CAPP polypeptide may include, but are not limited to those encoding the amino acid sequence of the mature polypeptide, by itself; the coding sequence for the mature polypeptide and additional sequences, such as those encoding the

300012: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

16

about 32 amino acid leader or secretory sequence; such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing, including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; an additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz *et al.*, *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The "HA" tag is another peptide useful for purification which corresponds to an epitope derived from the influenza hemagglutinin protein, which has been described by Wilson *et al.*, *Cell* 37: 767 (1984). As discussed below, other such fusion proteins include the CAPP fused to Fc at the N- or C-terminus.

The present invention further relates to variants of the nucleic acid molecules of the present invention, which encode portions, analogs or derivatives of the CAPP protein. Variants may occur naturally, such as a natural allelic variant. By an "allelic variant" is intended one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. *Genes II*, Lewin, B., ed., John Wiley & Sons, New York (1985). Non-naturally occurring variants may be produced using art-known mutagenesis techniques.

Such variants include those produced by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding regions, non-coding regions, or both. Alterations in the coding regions may produce conservative or

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

17

non-conservative amino acid substitutions, deletions or additions. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the CAPP protein or portions thereof. Also especially preferred in this regard are conservative substitutions. Most highly preferred are nucleic acid molecules encoding the mature protein having the amino acid sequence shown in Figure 1 (SEQ ID NO:2) or the mature CAPP amino acid sequence encoded by the deposited cDNA clone.

Further embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 95%, 96%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding the polypeptide having the amino acid sequence in SEQ ID NO:2; (b) a nucleotide sequence encoding the polypeptide having the amino acid sequence in SEQ ID NO:2, but lacking the N-terminal methionine; (c) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions from about 1 to about 365 in SEQ ID NO:2; (d) a nucleotide sequence encoding the polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; (e) a nucleotide sequence encoding the mature CAPP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; or (f) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), or (e).

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a CAPP polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the CAPP polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the

IDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

18

reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular nucleic acid molecule is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in Figure 1 or to the nucleotides sequence of the deposited cDNA clone can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711. Bestfit uses the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2: 482-489. (1981), to find the best segment of homology between two sequences. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference nucleotide sequence and that gaps in homology of up to 5% of the total number of nucleotides in the reference sequence are allowed.

The present application is directed to nucleic acid molecules at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) or to the nucleic acid sequence of the deposited cDNA, irrespective of whether they encode a polypeptide having CAPP activity. This is because even where a particular nucleic acid molecule does not encode a polypeptide having CAPP activity, one of skill in the art would still know how to use the nucleic acid molecule, for instance, as a hybridization probe or a polymerase chain reaction (PCR) primer. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having CAPP activity include, *inter alia*, (1) isolating the CAPP gene or allelic variants thereof in a cDNA library; (2) *in situ* hybridization (e.g., "FISH") to metaphase chromosomal

DOCID: AWO_8844112A1.J

WO 98/44112

PCT/US98/06022

19

spreads to provide precise chromosomal location of the CAPP gene, as described in Verma *et al.*, *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York (1988); and Northern Blot analysis for detecting CAPP mRNA expression in specific tissues.

5 Preferred, however, are nucleic acid molecules having sequences at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) or to the nucleic acid sequence of the deposited cDNA which do, in fact, encode a polypeptide having CAPP protein activity. By "a polypeptide having CAPP activity" is intended polypeptides exhibiting activity similar, but not necessarily identical, to an activity of the CAPP protein of the invention (either the full-length protein or, preferably, the mature protein), as measured in a particular biological assay.

10 Thus, "a polypeptide having CAPP protein activity" includes polypeptides that exhibit CAPP activity. Although the degree of activity need not be identical to that of the CAPP protein, preferably, "a polypeptide having CAPP protein activity" will exhibit substantially similar activity as compared to the CAPP protein (i.e., the candidate polypeptide will exhibit greater activity or not more than about twenty-fold less and, preferably, not more than about ten-fold less activity relative to the reference CAPP protein).

20 Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the nucleic acid molecules having a sequence at least 90%, 95%, 96%, 97%, 98% or 99% identical to the nucleic acid sequence of the deposited cDNA or the nucleic acid sequence shown in Figure 1 (SEQ ID NO:1) will encode a polypeptide "having CAPP protein activity." In fact, since degenerate variants of these nucleotide sequences all encode the same polypeptide, this will be clear to the skilled artisan even without performing the above described comparison assay. It will be further recognized in the art that, for such nucleic acid molecules that are not degenerate variants, a reasonable number will also encode a polypeptide having CAPP protein activity. This is because the skilled artisan is fully aware of amino acid substitutions that are either less likely or not likely to significantly effect protein

25

30

DOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

20

function (e.g., replacing one aliphatic amino acid with a second aliphatic amino acid).

5 For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selections or screens to identify sequences that maintain functionality. As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie, J.U. *et al.*, *supra*, and the references cited therein.

20 *Vectors and Host Cells*

The present invention also relates to vectors which include the isolated DNA molecules of the present invention, host cells which are genetically engineered with the recombinant vectors, and the production of CAPP polypeptides or fragments thereof by recombinant techniques.

25 Recombinant constructs may be introduced into host cells using well known techniques such infection, transduction, transfection, transvection, electroporation and transformation. The vector may be, for example, a phage, plasmid, viral or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

ISDOCID: <WO_9844112A_LL>

WO 98/44112

PCT/US98/06022

21

5 The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

Preferred are vectors comprising cis-acting control regions to the polynucleotide of interest. Appropriate trans-acting factors may be supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

10 In certain preferred embodiments in this regard, the vectors provide for specific expression, which may be inducible and/or cell type-specific. Particularly preferred among such vectors are those inducible by environmental factors that are easy to manipulate, such as temperature and nutrient additives.

15 Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors, e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episomes, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as cosmids and phagemids.

20 The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli lac*, *trp* and *tac* promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination and, 25 in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will preferably include a translation initiating at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

30 As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase or neomycin

ID00ID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

22

resistance for eukaryotic cell culture and tetracycline or ampicillin resistance genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

In addition to the use of expression vectors in the practice of the present invention, the present invention further includes novel expression vectors comprising operator and promoter elements operatively linked to nucleotide sequences encoding a protein of interest. One example of such a vector is pHE4a which is described in detail below.

As summarized in Figures 4 and 5, components of the pHE4a vector (SEQ ID NO: 8) include: 1) a neomycinphosphotransferase gene as a selection marker, 2) an *E. coli* origin of replication, 3) a T5 phage promoter sequence, 4) two *lac* operator sequences, 5) a Shine-Delgarno sequence, 6) the lactose operon repressor gene (*lacIq*) and 7) a multiple cloning site linker region. The origin of replication (*oriC*) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences were made synthetically. Synthetic production of nucleic acid sequences is well known in the art. CLONTECH 95/96 Catalog, pages 215-216, CLONTECH, 1020 East Meadow Circle, Palo Alto, CA 94303. The pHE4a vector was deposited with the ATCC (12301 Parklawn Drive, Rockville, Maryland 20852) on February, 1998, and given accession number 209645.

A nucleotide sequence encoding CAPP (SEQ ID NO:1), is operatively linked to the promoter and operator of pHE4a by restricting the vector with *NdeI* and either *XbaI*, *BamHI*, *XhoI*, or *Asp718*, and isolating the larger fragment (the multiple cloning site region is about 310 nucleotides) on a gel. The nucleotide sequence encoding CAPP (SEQ ID NO:1) having the appropriate restriction sites is generated, for example, according to the PCR protocol described in Example 1, using PCR primers having restriction sites for *NdeI* (as the 5' primer) and either *XbaI*, *BamHI*, *XhoI*, or *Asp718* (as the 3' primer). The PCR insert is gel purified

WO 98/44112

PCT/US98/06022

23

and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

As noted above, the pHE4a vector contains a *lacIq* gene. *LacIq* is an allele of the *lacI* gene which confers tight regulation of the *lac* operator. Amann, E. *et al.*, *Gene* 69:301-315 (1988); Stark, M., *Gene* 51:255-267 (1987). The *lacIq* gene encodes a repressor protein which binds to *lac* operator sequences and blocks transcription of down-stream (i.e., 3') sequences. However, the *lacIq* gene product dissociates from the *lac* operator in the presence of either lactose or certain lactose analogs, e.g., isopropyl B-D-thiogalactopyranoside (IPTG). CAPP thus is not produced in appreciable quantities in uninduced host cells containing the pHE4a vector. Induction of these host cells by the addition of an agent such as IPTG, however, results in the expression of the CAPP coding sequence.

The promoter/operator sequences of the pHE4a vector (SEQ ID NO:9) comprise a T5 phage promoter and two *lac* operator sequences. One operator is located 5' to the transcriptional start site and the other is located 3' to the same site. These operators, when present in combination with the *lacIq* gene product, confer tight repression of down-stream sequences in the absence of a *lac* operon inducer, e.g., IPTG. Expression of operatively linked sequences located down-stream from the *lac* operators may be induced by the addition of a *lac* operon inducer, such as IPTG. Binding of a *lac* inducer to the *lacIq* proteins results in their release from the *lac* operator sequences and the initiation of transcription of operatively linked sequences. *Lac* operon regulation of gene expression is reviewed in Devlin, T., TEXTBOOK OF BIOCHEMISTRY WITH CLINICAL CORRELATIONS, 4th Edition (1997), pages 802-807.

The pHE4 series of vectors contain all of the components of the pHE4a vector except for the CAPP coding sequence. Features of the pHE4a vectors include optimized synthetic T5 phage promoter, *lac* operator, and Shine-Delagarno sequences. Further, these sequences are also optimally spaced so that expression of an inserted gene may be tightly regulated and high level of expression occurs upon induction.

WO 98/44112

PCT/US98/06022

24

Among known bacterial promoters suitable for use in the production of proteins of the present invention include the *E. coli lacI* and *lacZ* promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I promoter.

The pHE4a vector also contains a Shine-Delgarno sequence 5' to the AUG initiation codon. Shine-Delgarno sequences are short sequences generally located about 10 nucleotides up-stream (*i.e.*, 5') from the AUG initiation codon. These sequences essentially direct prokaryotic ribosomes to the AUG initiation codon.

Thus, the present invention is also directed to expression vector useful for the production of the proteins of the present invention. This aspect of the invention is exemplified by the pHE4a vector (SEQ ID NO:8).

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Among known bacterial promoters suitable for use in the present invention include the *E. coli lacI* and *lacZ* promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I promoter.

WO 98/44112

PCT/US98/06022

25

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis *et al.*, *Basic Methods In Molecular Biology* (1986).

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals, but also additional heterologous functional regions. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification, or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize proteins. For example, EP A 0,464,533 (Canadian counterpart 2,045,869) discloses fusion proteins comprising

DOCID: WO_9844112A1_1

WO 98/44112

PCT/US98/06022

26

various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is thoroughly advantageous for use in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP A 0,232,262).

5 On the other hand, for some uses it would be desirable to be able to delete the Fc part after the fusion protein has been expressed, detected and purified in the advantageous manner described. This is the case when Fc portion proves to be a hindrance to use in therapy and diagnosis, for example when the fusion protein is to be used as antigen for immunizations. In drug discovery, for example,
10 human proteins, such as, hIL-5-receptor has been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. See, Bennett *et al.*, *Journal of Molecular Recognition* 8:52-58 (1995) and Johanson *et al.*, *J. Biol. Chem.* 270(16):9459-9471 (1995).

The CAPP protein can be recovered and purified from recombinant cell
15 cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography
20 ("HPLC") is employed for purification. Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant
25 production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

27

CAPP Polypeptides and Fragments

The invention further provides an isolated CAPP polypeptide having the amino acid sequence encoded by the deposited cDNA, or the amino acid sequence in Figure 1 (SEQ ID NO:2), or a peptide or polypeptide comprising a portion of the above polypeptides. The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least to amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than ten amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus.

It will be recognized in the art that some amino acid sequences of the CAPP polypeptide can be varied without significant effect of the structure or function of the protein. If such differences in sequence are contemplated, it should be remembered that there will be critical areas on the protein which determine activity. In general, it is possible to replace residues which form the tertiary structure, provided that residues performing a similar function are used. In other instances, the type of residue may be completely unimportant if the alteration occurs at a non-critical region of the protein.

Thus, the invention further includes variations of the CAPP polypeptide which show substantial CAPP polypeptide activity or which include regions of CAPP protein such as the protein portions discussed below. Such mutants include deletions, insertions, inversions, repeats, and type substitutions. As indicated above, guidance concerning which amino acid changes are likely to be phenotypically silent can be found in Bowie, J.U., *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," *Science* 247:1306-1310 (1990).

Thus, the fragment, derivative or analog of the polypeptide of SEQ ID NO:2, or that encoded by the deposited cDNA, may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-

3DOCID: <WD_9844112A1.L>

WO 98/44112

PCT/US98/06022

28

conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as an IgG Fc fusion region peptide or leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Of particular interest are substitutions of charged amino acids with another charged amino acid and with neutral or negatively charged amino acids. The latter results in proteins with reduced positive charge to improve the characteristics of the CAPP protein. The prevention of aggregation is highly desirable. Aggregation of proteins not only results in a loss of activity but can also be problematic when preparing pharmaceutical formulations, because they can be immunogenic. (Pinckard *et al.*, *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins *et al.*, *Diabetes* 36:838-845 (1987); Cleland *et al.*, *Crit. Rev. Therapeutic Drug Carrier Systems* 10:307-377 (1993)).

The replacement of amino acids can also change the selectivity of binding to cell surface receptors. Ostade *et al.*, *Nature* 361:266-268 (1993) describes certain mutations resulting in selective binding of TNF- α to only one of the two known types of TNF receptors. Thus, the CAPP of the present invention may include one or more amino acid substitutions, deletions or additions, either from natural mutations or human manipulation.

As indicated, changes are preferably of a minor nature, such as conservative amino acid substitutions that do not significantly affect the folding or activity of the protein (see Table 1).

WO 98/44112

PCT/US98/06022

29

TABLE 1. Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine Tryptophan Tyrosine
Hydrophobic	Leucine Isoleucine Valine
Polar	Glutamine Asparagine
Basic	Arginine Lysine Histidine
Acidic	Aspartic Acid Glutamic Acid
Small	Alanine Serine Threonine Methionine Glycine

Of course, the number of amino acid substitutions a skilled artisan would make depends on many factors, including those described above. Generally speaking, the number of amino acid substitutions for any given CAPP polypeptide will not be more than 50, 40, 30, 20, 10, 5, or 3.

Amino acids in the CAPP protein of the present invention that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992) and de Vos *et al.*, *Science* 255:306-312 (1992)).

S00CID: <WO_9844112A1_1_>

WO 98/44112

PCT/US98/06022

30

The polypeptides of the present invention are preferably provided in an isolated form. By "isolated polypeptide" is intended a polypeptide removed from its native environment. Thus, a polypeptide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" are polypeptides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of the CAPP polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

The polypeptides of the present invention include the polypeptide encoded by the deposited cDNA including the leader; the mature polypeptide encoded by the deposited the cDNA minus the leader (i.e., the mature protein); a polypeptide comprising amino acids about -32 to about 365 in SEQ ID NO:2; a polypeptide comprising amino acids about -31 to about 365 in SEQ ID NO:2; a polypeptide comprising amino acids about 1 to about 365 in SEQ ID NO:2; as well as polypeptides which have at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98% or 99% similarity to those described above. Further polypeptides of the present invention include polypeptides at least 80% identical, more preferably at least 90% or 95% identical, still more preferably at least 96%, 97%, 98% or 99% identical to the polypeptide encoded by the deposited cDNA, to the polypeptide of Figure 1 (SEQ ID NO:2), and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

By "% similarity" for two polypeptides is intended a similarity score produced by comparing the amino acid sequences of the two polypeptides using the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) and the default settings for determining similarity. Bestfit uses the local homology algorithm of Smith and Waterman (*Advances in Applied Mathematics* 2: 482-489, 1981) to find the best segment of similarity between two sequences.

SDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

31

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a reference amino acid sequence of a CAPP polypeptide is intended that the amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid of the CAPP polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequence shown in Figure 1 (SEQ ID NO:2) or to the amino acid sequence encoded by deposited cDNA clone can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711. When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence according to the present invention, the parameters are set, of course, such that the percentage of identity is calculated over the full length of the reference amino acid sequence and that gaps in homology of up to 5% of the total number of amino acid residues in the reference sequence are allowed.

The polypeptide of the present invention could be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art.

DOCID: WO_9844112A1_1

WO 98/44112

PCT/US98/06022

32

As described in detail below, the polypeptides of the present invention can also be used to raise polyclonal and monoclonal antibodies, which are useful in assays for detecting CAPP protein expression as described below or as agonists and antagonists capable of enhancing or inhibiting CAPP protein function. Further, such polypeptides can be used in the yeast two-hybrid system to "capture" CAPP protein binding proteins which are also candidate agonist and antagonist according to the present invention. The yeast two hybrid system is described in Fields and Song, *Nature* 340:245-246 (1989).

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention. The epitope of this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. These immunogenic epitopes are believed to be confined to a few loci on the molecule. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1983).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind), it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G. *et al.* "Antibodies that react with predetermined sites on proteins," *Science* 219:660-666 (1983). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the

SDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

33

mimicked protein; longer, peptides, especially those containing proline residues, usually are effective. Sutcliffe *et al.*, *supra*, at 661. For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HA1 polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the rabies glycoprotein induced antibodies that precipitated the respective proteins.

Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. Thus, a high proportion of hybridomas obtained by fusion of spleen cells from donors immunized with an antigen epitope-bearing peptide generally secrete antibody reactive with the native protein. Sutcliffe *et al.*, *supra*, at 663. The antibodies raised by antigenic epitope-bearing peptides or polypeptides are useful to detect the mimicked protein, and antibodies to different peptides may be used for tracking the fate of various regions of a protein precursor which undergoes post-translational processing. The peptides and anti-peptide antibodies may be used in a variety of qualitative or quantitative assays for the mimicked protein, for instance in competition assays since it has been shown that even short peptides (e.g., about 9 amino acids) can bind and displace the larger peptides in immunoprecipitation assays. See, for instance, Wilson *et al.*, *Cell* 37:767-778 (1984) at 777. The anti-peptide antibodies of the invention also are useful for purification of the mimicked protein, for instance, by adsorption chromatography using methods well known in the art.

Antigenic epitope-bearing peptides and polypeptides of the invention designed according to the above guidelines preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of a polypeptide of the invention, containing about 30 to about 50 amino acids, or any length up to and including

SDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

34

the entire amino acid sequence of a polypeptide of the invention, also are considered epitope-bearing peptides or polypeptides of the invention and also are useful for inducing antibodies that react with the mimicked protein. Preferably, the amino acid sequence of the epitope-bearing peptide is selected to provide
5 substantial solubility in aqueous solvents (i.e., the sequence includes relatively hydrophilic residues and highly hydrophobic sequences are preferably avoided); and sequences containing proline residues are particularly preferred.

Non-limiting examples of antigenic polypeptides or peptides that can be used to generate CAPP-specific antibodies include: a polypeptide comprising
10 amino acid residues from about -32 to about -22 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about -4 to about 40 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 46 to about 57 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 62 to about 73 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about
15 78 to about 87 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 92 to about 110 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 119 to about 144 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 152 to about 186 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 200 to about 219 in
20 SEQ ID NO:2; a polypeptide comprising amino acid residues from about 230 to about 240 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 248 to about 258 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 314 to about 336 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 344 to about 353 in SEQ ID NO:2.
25 As indicated above, the inventors have determined that the above polypeptide fragments are antigenic regions of the CAPP protein.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means for making peptides or polypeptides including recombinant means using nucleic acid molecules of the invention. For
30 instance, a short epitope-bearing amino acid sequence may be fused to a larger polypeptide which acts as a carrier during recombinant production and

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

35

5 purification, as well as during immunization to produce anti-peptide antibodies. Epitope-bearing peptides also may be synthesized using known methods of chemical synthesis. For instance, Houghten has described a simple method for synthesis of large numbers of peptides, such as 10-20 mg of 248 different 13
10 residue peptides representing single amino acid variants of a segment of the HA1 polypeptide which were prepared and characterized (by ELISA-type binding studies) in less than four weeks. Houghten, R. A., "General method for the rapid solid-phase synthesis of large numbers of peptides: specificity of antigen-antibody interaction at the level of individual amino acids," *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further described in U.S. Patent No. 4,631,211 to Houghten *et al.* (1986). In this procedure the individual resins for the solid-phase synthesis of various peptides are contained in separate solvent-permeable packets, enabling the optimal use of the many identical repetitive steps involved in
15 solid-phase methods. A completely manual procedure allows 500-1000 or more syntheses to be conducted simultaneously. Houghten *et al.*, *supra*, at 5134.

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe *et al.*, *supra*; Wilson *et al.*, *supra*; Chow, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J. *et al.*, *J. Gen. Virol.* 66:2347-2354 (1985).
20 Generally, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemacyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine may be coupled to carrier using a linker
25 such as m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carrier using a more general linking agent such as glutaraldehyde. Animals such as rabbits, rats and mice are immunized with either free or carrier-coupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 mg peptide or carrier
30 protein and Freund's adjuvant. Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide

DOCID: WO_9844112A1_1

WO 98/44112

PCT/US98/06022

36

antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid support and elution of the selected antibodies according to methods well known in the art.

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. For instance, Geysen *et al.*, *supra*, discloses a procedure for rapid concurrent synthesis on solid supports of hundreds of peptides of sufficient purity to react in an enzyme-linked immunosorbent assay. Interaction of synthesized peptides with antibodies is then easily detected without removing them from the support. In this manner a peptide bearing an immunogenic epitope of a desired protein may be identified routinely by one of ordinary skill in the art. For instance, the immunologically important epitope in the coat protein of foot-and-mouth disease virus was located by Geysen *et al.* with a resolution of seven amino acids by synthesis of an overlapping set of all 208 possible hexapeptides covering the entire 213 amino acid sequence of the protein. Then, a complete replacement set of peptides in which all 20 amino acids were substituted in turn at every position within the epitope were synthesized, and the particular amino acids conferring specificity for the reaction with antibody were determined. Thus, peptide analogs of the epitope-bearing peptides of the invention can be made routinely by this method. U.S. Patent No. 4,708,781 to Geysen (1987) further describes this method of identifying a peptide bearing an immunogenic epitope of a desired protein.

Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is

WO 98/44112

PCT/US98/06022

37

complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. *et al.* (1996) on "Peralkylated Oligopeptide Mixtures" discloses linear C₁-C₇-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

The entire disclosure of each document cited in this section on "Polypeptides and Peptides" is hereby incorporated herein by reference.

As one of skill in the art will appreciate, CAPP polypeptides of the present invention and the epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life *in vivo*. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EP A 394,827; Traunecker *et al.*, *Nature* 331:84- 86 (1988)). Fusion proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric CAPP protein or protein fragment alone (Fountoulakis *et al.*, *J Biochem* 270:3958-3964 (1995)).

Disease State Diagnosis and Prognosis

It is believed that certain maladies in mammals may cause the mammals to express significantly altered levels of the CAPP protein and mRNA encoding the CAPP protein when compared to a corresponding "standard" mammal, i.e., a mammal of the same species not having the malady or condition. For example, a mammal suffering from pancreatitis or a condition that causes abnormal myocardial hypertrophy is expected to express altered levels of CAPP by the pancreas or heart, respectively. Further, it is believed that decreased levels of the

DOCID: WO_9844112A1_1

WO 98/44112

PCT/US98/06022

38

CAPP protein can be detected in certain body fluids (e.g., sera, plasma, urine, and spinal fluid) from mammals with such a condition when compared to sera from mammals of the same species not having the condition. Thus, the invention provides a diagnostic method useful during diagnosis of pancreatitis or one of the many conditions that cause abnormal hypertrophy of the heart, such as hypertension, myocardial infarction, valve disease and cardiomyopathy. The method involves assaying the expression level of the gene encoding the CAPP protein in mammalian cells or body fluid and comparing the gene expression level with a standard CAPP gene expression level, whereby a decrease in the gene expression level over the standard is indicative of said conditions.

Where a diagnosis has already been made according to conventional methods, the present invention is useful as a prognostic indicator, whereby patients exhibiting decreased CAPP gene expression will experience a worse clinical outcome relative to patients expressing the gene at a lower level.

Additionally, the presence of CAPP protein or mRNA level can be measured to qualitatively determine cell or tissue type. Since CAPP is highly expressed in mature heart, pancreas and placenta tissue, CAPP expression can be employed to determine the type of cells that are present in a cell culture.

The CAPP gene was discovered in an activated T-cell cDNA library. CAPP protein and mRNA expression can be used as a marker to detect activated T-cells. Monitoring T cells activation is useful for a number of *in vitro* diagnostic purposes, including studying the effects of candidate drugs on the immune system, and determining whether the T cells of a subject have been activated by analyzing a blood sample taken from the subject or by assessing activity in an *in vitro* screening test.

By "assaying the expression level of the gene encoding the CAPP protein" is intended qualitatively or quantitatively measuring or estimating the level of the CAPP protein or the level of the mRNA encoding the CAPP protein in a first biological sample either directly (e.g., by determining or estimating absolute protein level or mRNA level) or relatively (e.g., by comparing to the CAPP protein level or mRNA level in a second biological sample).

ISDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

39

5 Preferably, the CAPP protein level or mRNA level in the first biological sample is measured or estimated and compared to a standard CAPP protein level or mRNA level, the standard being taken from a second biological sample obtained from an individual not having the cancer. As will be appreciated in the art, once a standard CAPP protein level or mRNA level is known, it can be used repeatedly as a standard for comparison.

10 By "biological sample" is intended any biological sample obtained from an individual, cell line, tissue culture, or other source which contains CAPP protein or mRNA. Biological samples include mammalian body fluids (such as sera, plasma, urine, synovial fluid and spinal fluid) which contain secreted mature CAPP protein, and heart, placenta, pancreas and umbilical tissue. Methods for obtaining tissue biopsies and body fluids from mammals are well known in the art. Where the biological sample is to include mRNA, a tissue biopsy is the preferred source.

15 Preferred mammals include monkeys, apes, cats, dogs, cows, pigs, horses, rabbits and humans. Particularly preferred are humans.

20 Total cellular RNA can be isolated from a biological sample using any suitable technique such as the single-step guanidinium-thiocyanate-phenol-chloroform method described in Chomczynski and Sacchi, *Anal. Biochem.* 162:156-159 (1987). Levels of mRNA encoding the CAPP protein are then assayed using any appropriate method. These include Northern blot analysis, S1 nuclease-mapping, the polymerase chain reaction (PCR), reverse transcription in combination with the polymerase chain reaction (RT-PCR); and reverse transcription in combination with the ligase chain reaction (RT-LCR).

25 Northern blot analysis can be performed as described in Harada *et al.*, *Cell* 63:303-312 (1990). Briefly, total RNA is prepared from a biological sample as described above. For the Northern blot, the RNA is denatured in an appropriate buffer (such as glyoxal/dimethyl sulfoxide/sodium phosphate buffer), subjected to agarose gel electrophoresis, and transferred onto a nitrocellulose filter. After 30 the RNAs have been linked to the filter by a UV linker, the filter is prehybridized in a solution containing formamide, SSC, Denhardt's solution, denatured salmon

300010: <WO 984112A1>

WO 98/44112

PCT/US98/06022

40

5 sperm, SDS, and sodium phosphate buffer. CAPP protein cDNA labeled according to any appropriate method (such as the ^{32}P -multiprimered DNA labeling system (Amersham)) is used as probe. After hybridization overnight, the filter is washed and exposed to x-ray film. cDNA for use as probe according to the present invention is described in the sections above and will preferably at least 15 bp in length.

10 S1 mapping can be performed as described in Fujita *et al.*, *Cell* 49:357-367 (1987). To prepare probe DNA for use in S1 mapping, the sense strand of above-described cDNA is used as a template to synthesize labeled antisense DNA. The antisense DNA can then be digested using an appropriate restriction endonuclease to generate further DNA probes of a desired length. Such antisense probes are useful for visualizing protected bands corresponding to the target mRNA (i.e., mRNA encoding the CAPP protein). Northern blot analysis can be performed as described above.

15 Preferably, levels of mRNA encoding the CAPP protein are assayed using the RT-PCR method described in Makino *et al.*, *Technique* 2:295-301 (1990). By this method, the radioactivities of the "amplicons" in the polyacrylamide gel bands are linearly related to the initial concentration of the target mRNA. Briefly, this method involves adding total RNA isolated from a biological sample in a reaction mixture containing a RT primer and appropriate buffer. After incubating for primer annealing, the mixture can be supplemented with a RT buffer, dNTPs, DTT, RNase inhibitor and reverse transcriptase. After incubation to achieve reverse transcription of the RNA, the RT products are then subject to PCR using labeled primers. Alternatively, rather than labeling the primers, a labeled dNTP can be included in the PCR reaction mixture. PCR amplification can be performed in a DNA thermal cycler according to conventional techniques. After a suitable number of rounds to achieve amplification, the PCR reaction mixture is electrophoresed on a polyacrylamide gel. After drying the gel, the radioactivity of the appropriate bands (corresponding to the mRNA encoding the CAPP protein)) is quantified using an imaging analyzer. RT and PCR reaction ingredients and conditions, reagent and gel concentrations, and labeling methods

20

25

30

SDOCID: <WO_0044112A1_>

WO 98/44112

PCT/US98/06022

41

are well known in the art. Variations on the RT-PCR method will be apparent to the skilled artisan.

Any set of oligonucleotide primers which will amplify reverse transcribed target mRNA can be used and can be designed as described in the sections above.

5 Assaying CAPP protein levels in a biological sample can occur using any art-known method. Preferred for assaying CAPP protein levels in a biological sample are antibody-based techniques. For example, CAPP protein expression in tissues can be studied with classical immunohistological methods. In these, the specific recognition is provided by the primary antibody (polyclonal or
10 monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunohistological staining of tissue section for pathological examination is obtained. Tissues can also be extracted, e.g., with urea and neutral detergent, for the liberation of CAPP protein for Western-blot or dot/slot assay (Jalkanen, M., *et al.*, *J. Cell. Biol.* 101:976-985 (1985); Jalkanen, M., *et al.*, *J. Cell. Biol.* 105:3087-3096 (1987)).
15 In this technique, which is based on the use of cationic solid phases, quantitation of CAPP protein can be accomplished using isolated CAPP protein as a standard. This technique can also be applied to body fluids. With these samples, a molar concentration of CAPP protein will aid to set standard values of CAPP protein content for different body fluids, like serum, plasma, urine, spinal fluid, etc. The
20 normal appearance of CAPP protein amounts can then be set using values from healthy individuals, which can be compared to those obtained from a test subject.

Other antibody-based methods useful for detecting CAPP protein gene expression include immunoassays, such as the enzyme linked immunosorbent
25 assay (ELISA) and the radioimmunoassay (RIA). For example, a CAPP protein-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe to detect and quantify the CAPP protein. The amount of CAPP protein present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer
30 algorithm. Such an ELISA for detecting a tumor antigen is described in Iacobelli *et al.*, *Breast Cancer Research and Treatment* 11:19-30 (1988). In another

DOCID: <WO_984112A1_1>

WO 98/44112

PCT/US98/06022

42

ELISA assay, two distinct specific monoclonal antibodies can be used to detect CAPP protein in a body fluid. In this assay, one of the antibodies is used as the immunoabsorbent and the other as the enzyme-labeled probe.

5 The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting CAPP protein with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the
10 assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample.

Suitable enzyme labels include, for example, those from the oxidase group, which catalyze the production of hydrogen peroxide by reacting with substrate. Glucose oxidase is particularly preferred as it has good stability and its
15 substrate (glucose) is readily available. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labelled antibody/substrate reaction. Besides enzymes, other suitable labels include radioisotopes, such as iodine (^{125}I , ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^3H), indium (^{111}In), and technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.
20

In addition to assaying CAPP protein levels in a biological sample obtained from an individual, CAPP protein can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of CAPP protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable
25 labels include radioisotopes such as barium or caesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

30 A CAPP protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for

ISOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

43

example, ^{131}I , ^{111}In , $^{99\text{m}}\text{Tc}$), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously or intraperitoneally) into the mammal to be examined for cancer. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of $^{99\text{m}}\text{Tc}$. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain CAPP protein. *In vivo* tumor imaging is described in S.W. Burchiel *et al.*, "Immunopharmacokinetics of Radiolabelled Antibodies and Their Fragments" (Chapter 13 in *Tumor Imaging: The Radiochemical Detection of Cancer*, eds., S.W. Burchiel and B.A. Rhodes, Masson Publishing Inc. (1982)).

CAPP-protein specific antibodies for use in the present invention can be raised against the intact CAPP protein or an antigenic polypeptide fragment thereof, which may presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')_2 fragments) which are capable of specifically binding to CAPP protein. Fab and F(ab')_2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl *et al.*, *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the CAPP protein or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of CAPP protein is prepared and purified to render it substantially

WO 98/44112

PCT/US98/06022

44

free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

5 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or CAPP protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology (Kohler *et al.*, *Nature* 256:495 (1975); Kohler *et al.*, *Eur. J. Immunol.* 6:511 (1976); Kohler *et al.*, *Eur. J. Immunol.* 6:292 (1976); Hammerling *et al.*, In: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981)). In general, such procedures involve immunizing an animal (preferably a mouse) with a CAPP protein antigen or, more preferably, with a CAPP protein-expressing cell. Suitable cells can be recognized by their capacity to bind anti-CAPP protein antibody. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 g/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP₂O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.* (*Gastroenterology* 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the CAPP protein antigen.

25 Alternatively, additional antibodies capable of binding to the CAPP protein antigen may be produced in a two-step procedure through the use of anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and that, therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, CAPP-protein specific antibodies are used to immunize an animal, preferably a mouse. The

WO 98/44112

PCT/US98/06022

45

5 splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the CAPP protein-specific antibody can be blocked by the CAPP protein antigen. Such antibodies comprise anti-idiotypic antibodies to the CAPP protein-specific antibody and can be used to immunize an animal to induce formation of further CAPP protein-specific antibodies.

10 It will be appreciated that Fab and F(ab')₂ and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). Alternatively, CAPP protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

15 Where *in vivo* imaging is used to detect enhanced levels of CAPP protein for tumor diagnosis in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. See, for review, Morrison, *Science* 229:1202 (1985); Oi *et al.*, *BioTechniques* 4:214 (1986); Cabilly *et al.*, U.S. Patent No. 4,816,567; Taniguchi *et al.*, EP 171496; Morrison *et al.*, EP 173494; Neuberger *et al.*, WO 8601533; Robinson *et al.*, WO 20 8702671; Boulianne *et al.*, *Nature* 312:643 (1984); Neuberger *et al.*, *Nature* 314:268 (1985).

25 Further suitable labels for the CAPP protein-specific antibodies of the present invention are provided below. Examples of suitable enzyme labels include malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast-alcohol dehydrogenase, alpha-glycerol phosphate dehydrogenase, triose phosphate isomerase, peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, 30 glucose-6-phosphate dehydrogenase, glucoamylase, and acetylcholine esterase.

DOCID: AWO_9844112A1_1_2

WO 98/44112

PCT/US98/06022

46

5 Examples of suitable radioisotopic labels include ^3H , ^{111}In , ^{125}I , ^{131}I , ^{32}P , ^{35}S , ^{14}C , ^{51}Cr , ^{57}Co , ^{58}Co , ^{59}Fe , ^{75}Se , ^{152}Eu , ^{90}Y , ^{67}Cu , ^{212}Bi , ^{211}At , ^{212}Pb , ^{47}Sc , ^{109}Pd , $^{99\text{m}}\text{Tc}$, etc. ^{111}In and $^{99\text{m}}\text{Tc}$ are preferred isotopes where *in vivo* imaging is used since it avoids the problem of dehalogenation of the ^{125}I or ^{131}I -labeled monoclonal antibody by the liver. In addition, this radionuclide has a more favorable gamma emission energy for imaging (Perkins *et al.*, *Eur. J. Nucl. Med.* 10:296-301 (1985); Carasquillo *et al.*, *J. Nucl. Med.* 28:281-287 (1987)). For example, ^{111}In coupled to monoclonal antibodies with 1-(P-isothiocyanatobenzyl)-DPTA has shown little uptake in non-tumorous tissues, particularly the liver, and therefore enhances specificity of tumor localization (Esteban *et al.*, *J. Nucl. Med.* 28:861-870 (1987)).

10 Examples of suitable non-radioactive isotopic labels include ^{157}Gd , ^{55}Mn , ^{162}Dy , ^{52}Tr , and ^{56}Fe , preferably chelated with a complexing ligand suitable for *in vivo* use.

15 Examples of suitable fluorescent labels include an ^{152}Eu label, a fluorescein label, an isothiocyanate label, a rhodamine label, a phycoerythrin label, a phycocyanin label, an allophycocyanin label, an o-phthaldehyde label, and a fluorescamine label.

20 Examples of suitable toxin labels include diphtheria toxin, ricin, and cholera toxin.

Examples of chemiluminescent labels include a luminal label, an isoluminal label, an aromatic acridinium ester label, an imidazole label, an acridinium salt label, an oxalate ester label, a luciferin label, a luciferase label, and an aequorin label.

25 Examples of nuclear magnetic resonance contrasting agents include heavy metal nuclei such as Gd, Mn, and iron, preferably chelated with a complexing ligand suitable for *in vivo* use.

30 Typical techniques for binding the above-described labels to antibodies are provided by Kennedy *et al.*, *Clin. Chim. Acta* 70:1-31 (1976), and Schurs *et al.*, *Clin. Chim. Acta* 81:1-40 (1977). Coupling techniques mentioned in the latter are the glutaraldehyde method, the periodate method, the dimaleimide method,

WO 98/44112

PCT/US98/06022

47

the *m*-maleimidobenzyl-N-hydroxy-succinimide ester method, all of which methods are incorporated by reference herein.

Therapeutics and Use in Cell Culture

5 The inventors contemplate that the CAPP polypeptide functions as a growth factor or similar cellular signaling polypeptide *in vivo*. CAPP possesses homology to the *Drosophila* brainiac polypeptide. See, Figure 2. This polypeptide is a neurogenic secreted molecule that is believed to play a role in the differentiation of embryonic cells into neurons. Thus, it is contemplated that the CAPP polypeptide exerts an effect on the differentiation of cells in the early stages of cell and tissue development, and may serve to aid in the differentiation of embryonic cells into heart or pancreas cells.

10 The CAPP polypeptide is also highly expressed in adult heart and pancreas tissue. One role of CAPP in mature muscle tissue may be to inhibit cell replication and division in the mature muscle tissue.

15 Thus, the inventors contemplate a number of additional practical utilities that use the growth-affecting properties of the CAPP polypeptide to modulate the differentiation and proliferation of cells and tissue, both *in vivo* and *ex vivo*.

Assessing the modulating effects of the CAPP polypeptide on the cellular proliferation and differentiation of cells can be performed as described below.

20 Biological activity of CAPP polypeptides can be examined in organ culture assays or in colony assay systems in agarose culture. Stimulation or inhibition of cellular proliferation may be measured by a variety of assays. For observing cell growth inhibition, one can use a solid or liquid medium. In a solid medium, cells undergoing growth inhibition can easily be selected from the subject cell group by comparing the sizes of colonies formed. In a liquid medium, growth inhibition can be screened by measuring culture broth turbidity or incorporation of labeled thymidine into DNA. Typically, the incorporation of a nucleoside analog into newly synthesized DNA is employed to measure proliferation (active cell growth) in a population of cells. For example, bromodeoxyuridine (BrdU) can be

25 employed as a DNA labeling reagent and Anti-BrdU mouse monoclonal antibody

30

SDOQID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

48

(clone BMC 9318 IgG₁) can be employed as a detection reagent. This antibody binds only to cells containing DNA which has incorporated bromodeoxyuridine. A number of detection methods can be used in conjunction with this assay including immunofluorescence, immunohistochemical, ELISA and colorimetric methods. Kits that include bromodeoxyuridine (BrdU) and Anti-BrdU mouse monoclonal antibody are commercially available from Boehringer Mannheim (Indianapolis, IN).

Effect upon cellular differentiation can be measured by contacting embryonic cells with various amounts of a CAPP polypeptide and observing the effect upon differentiation of the embryonic cells. Tissue specific antibodies and microscopy may be used to identify the resulting cells.

Modes of administration

It will be appreciated that conditions caused by a decrease in the standard or normal level of CAPP activity in an individual, can be treated by administration of CAPP protein. Thus, the invention further provides a method of treating an individual in need of an increased level of CAPP activity comprising administering to such an individual a pharmaceutical composition comprising an effective amount of an isolated CAPP polypeptide of the invention, particularly a mature form of the CAPP, effective to increase the CAPP activity level in such an individual.

The CAPP polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with CAPP polypeptide alone), the site of delivery of the CAPP polypeptide composition, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" of CAPP polypeptide for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of CAPP polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted

SDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

49

above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the CAPP polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 μ g/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the CAPP of the invention may be administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, drops or transdermal patch), buccally, or as an oral or nasal spray. By "pharmaceutically acceptable carrier" is meant a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The CAPP polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. *et al.*, *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer *et al.*, *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer *et al.*, *Id.*) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release CAPP polypeptide compositions also include liposomally entrapped CAPP polypeptide. Liposomes containing CAPP polypeptide are prepared by methods known *per se*: DE 3,218,121; Epstein *et al.*, *Proc. Natl. Acad. Sci. (USA)* 82:3688-3692 (1985); Hwang *et al.*, *Proc. Natl. Acad. Sci. (USA)* 77:4030-4034

300010: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

50

(1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal CAPP polypeptide therapy.

For parenteral administration, in one embodiment, the CAPP polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the CAPP polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrans; chelating agents such

SDOCID: <WO_9844112A1_L>

WO 98/44112

PCT/US98/06022

51

as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

5 The CAPP polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of CAPP polypeptide salts.

10 CAPP polypeptide to be used for therapeutic administration must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic CAPP polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

15 CAPP polypeptide ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous CAPP polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized CAPP polypeptide using bacteriostatic Water-for-Injection.

20 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

25 For use in cell culture media, the CAPP polypeptide can be added to a culture medium to aid in the differentiation and maintenance of cultured heart,

WO 98/44112

PCT/US98/06022

52

pancreas and placenta cells. Useful concentration ranges for this purpose are from about 10 picograms/mL to about 10 micrograms/mL.

Chromosome Assays

5 The nucleic acid molecules of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal
10 location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

In certain preferred embodiments in this regard, the cDNA herein disclosed is used to clone genomic DNA of a CAPP protein gene. This can be
15 accomplished using a variety of well known techniques and libraries, which generally are available commercially. The genomic DNA then is used for *in situ* chromosome mapping using well known techniques for this purpose. Typically, in accordance with routine procedures for chromosome mapping, some trial and error may be necessary to identify a genomic probe that gives a good *in situ*
20 hybridization signal.

In addition, in some cases, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the
25 amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified portion.

30 PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with

WO 98/44112

PCT/US98/06022

53

the same oligonucleotide primers, sublocalization can be achieved with panels of portions from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization ("FISH") of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with probes from the cDNA as short as 50 or 60 bp. For a review of this technique, see Verma *et al.*, *Human Chromosomes: A Manual Of Basic Techniques*, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, *Mendelian Inheritance In Man*, available on-line through Johns Hopkins University, Welch Medical Library. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. This assumes 1 megabase mapping resolution and one gene per 20 kb.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

WO 98/44112

PCT/US98/06022

54

Examples

Example 1: Expression and Purification of CAPP protein in *E. coli*

5 The DNA sequence encoding the mature CAPP protein in the deposited cDNA clone is amplified using PCR oligonucleotide primers specific to the amino terminal sequences of the CAPP protein and to vector sequences 3' to the gene. Additional nucleotides containing restriction sites to facilitate cloning are added to the 5' and 3' sequences respectively.

The 5' oligonucleotide primer has the sequence:

10 5' AGCA GGATCC CAA GAA AAA AAT GGA AAA GGG 3' (SEQ ID NO:4) containing the underlined BamHI restriction site, which encodes 21 nucleotides of coding sequence in Figure 1 (SEQ ID NO:1) beginning with Q after the S in the signal peptide.

The 3' primer has the sequence:

15 5' ATTG TCTAGA TAT CTA TTT TAG CAT TTT A 3' (SEQ ID NO:5) containing the underlined XbaI restriction site followed by 19 nucleotides of sequence, including the last 8 nucleotides of the coding domain in Figure 1.

20 The restriction sites are convenient to restriction enzyme sites in the bacterial expression vector pQE9, which are used for bacterial expression in these examples. (Qiagen, Inc. 9259 Eton Avenue, Chatsworth, CA, 91311). pQE9 encodes ampicillin antibiotic resistance ("Amp^r") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), a 6-His tag and restriction enzyme sites.

25 The amplified CAPP DNA and the vector pQE9 both are digested with BamHI and XbaI and the digested DNAs are then ligated together. Insertion of the CAPP protein DNA into the restricted pQE9 vector places the CAPP protein coding region downstream of and operably linked to the vector's IPTG-inducible promoter and in-frame with an initiating AUG appropriately positioned for translation of CAPP protein.

30 The ligation mixture is transformed into competent *E. coli* cells using standard procedures. Such procedures are described in Sambrook *et al.*,

WO 98/44112

PCT/US98/06022

55

Molecular Cloning: a Laboratory Manual, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). *E. coli* strain MR15/rep4, containing multiple copies of the plasmid pREP4, which expresses lac repressor and confers kanamycin resistance ("Kan^r"), is used in carrying out the illustrative example described herein. This strain, which is only one of many that are suitable for expressing CAPP protein, is available commercially from Qiagen.

Transformants are identified by their ability to grow on LB plates in the presence of ampicillin and kanamycin. Plasmid DNA is isolated from resistant colonies and the identity of the cloned DNA confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 µg/ml) and kanamycin (25 µg/ml).

The O/N culture is used to inoculate a large culture, at a dilution of approximately 1:100 to 1:250. The cells are grown to an optical density at 600 nm ("OD₆₀₀") of between 0.4 and 0.6. Isopropyl-B-D-thiogalactopyranoside ("IPTG") is then added to a final concentration of 1 mM to induce transcription from lac repressor sensitive promoters, by inactivating the lacI repressor. Cells subsequently are incubated further for 3 to 4 hours. Cells then are harvested by centrifugation and disrupted, by standard methods. Inclusion bodies are purified from the disrupted cells using routine collection techniques, and protein is solubilized from the inclusion bodies into 8M urea. The 8M urea solution containing the solubilized protein is passed over a PD-10 column in 2 x phosphate-buffered saline ("PBS"), thereby removing the urea, exchanging the buffer and refolding the protein. The protein is purified by a further step of chromatography to remove endotoxin. Then, it is sterile filtered. The sterile filtered protein preparation is stored in 2 x PBS at a concentration of 95 µ/ml.

Example 2: Cloning and Expression of CAPP protein in a Baculovirus Expression System

The DNA sequence encoding the full length human CAPP protein, ATCC # 97729, is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene as follows.

WO 98/44112

PCT/US98/06022

56

The 5' primer has the sequence:

5' GGCC GGATCC GCC ATC ATG AGT GTT GGA CGT CGA AGA AT 3'
(SEQ ID NO:6) containing the underlined BamHI restriction enzyme site followed by 6 nucleotides resembling an efficient signal for the initiation of translation in eukaryotic cells (Kozak, M., *J. Mol. Biol.* 196:947-950 (1987)), followed by 23 nucleotides of coding sequence of the human CAPP gene (the initiation codon for translation "ATG" is double underlined).

The 3' primer has the sequence:

5' ATTG TCTAGA TAT CTA TTT TAG CAT TTT A 3' (SEQ ID NO:5)
containing the underlined XbaI restriction site followed by 19 nucleotides of CAPP sequence, including the last 8 nucleotides of the coding domain.

The amplified sequences are isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment is then digested with the endonucleases BamHI and XbaI and then purified again on a 1% agarose gel. This fragment is designated F2.

The vector A2 (modification of pVL941 vector, discussed below) is used for expression of the human CAPP protein using the baculovirus expression system (see, Summers, M.D. and Smith, G.E. 1987, *a Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Agricultural Experimental Station Bulletin No. 1555). This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhidrosis virus (AcMNPV) followed by the recognition sites for the restriction endonucleases BamHI and Asp718. The polyadenylation site of the simian virus (SV)40 is used for efficient polyadenylation. For an easy selection of recombinant viruses the beta-galactosidase gene from *E.coli* is inserted in the same orientation as the polyhedrin promoter followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked on both sides by viral sequences for the cell-mediated homologous recombination of cotransfected wild-type viral DNA. Many other baculovirus vectors can be used in place of A2 such as pAc373, pVL941 and pAcIM1 (Luckow, V.A. and Summers, M.D., *Virology*, 170:31-39).

WO 98/44112

PCT/US98/06022

57

The plasmid is digested with the restriction enzymes XbaI and BamHI and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The DNA is then isolated from a 1% agarose gel using the commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated V2.

Fragment F2 and the dephosphorylated plasmid V2 are ligated with T4 DNA ligase. *E.coli* HB101 cells are then transformed and bacteria identified that contained the plasmid (pBac CAPP) with the CAPP gene using the PCR method, in which one of the primers is that used to amplify the gene and the second primer is from well within the vector so that only those bacterial colonies containing the CAPP gene fragment will show amplification of the DNA. The sequence of the cloned fragment is confirmed by DNA sequencing.

5 ug of the plasmid pBac CAPP is cotransfected with 1 ug of a commercially available linearized baculovirus ("BaculoGold baculovirus DNA", Pharmingen, San Diego, CA.) using the lipofection method (Felgner *et al. Proc. Natl. Acad. Sci. USA*, 84:7413-7417 (1987)).

1 ug of BaculoGold virus DNA and 5 ug of the plasmid pBac CAPP are mixed in a sterile well of a microtiter plate containing 50 μ l of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added dropwise to the Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is rocked back and forth to mix the newly added solution. The plate is then incubated for 5 hours at 27°C. After 5 hours the transfection solution is removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. The plate is put back into an incubator and cultivation continued at 27°C for four days.

After four days the supernatant is collected and a plaque assay performed similar to that described by Summers and Smith (*supra*). As a modification, an agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used which allows an easy isolation of blue stained plaques. (A detailed description of a

WO 98/44112

PCT/US98/06022

58

"plaque assay" can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after a serial dilution of the viruses is added to the cells, blue stained plaques are picked with the tip of an Pasteur pipette. The agar containing the recombinant viruses is then resuspended in an Eppendorf tube containing 200 μ l of Grace's medium. The agar is removed by a brief centrifugation and the supernatant containing the recombinant baculoviruses (designated baculovirus V-CAPP) is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then stored at 4°C.

Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus V-CAPP at a multiplicity of infection (MOI) of 2. Six hours later the medium is removed and replaced with SF900 II medium minus methionine and cysteine (Life Technologies Inc., Gaithersburg). 42 hours later 5 μ Ci of 35 S-methionine and 5 μ Ci 35 S cysteine (Amersham) are added. The cells are further incubated for 16 hours before they are harvested by centrifugation and the labelled proteins visualized by SDS-PAGE and autoradiography.

Example 3: Cloning and Expression in Mammalian Cells

Most of the vectors used for the transient expression of the CAPP protein gene sequence in mammalian cells should carry the SV40 origin of replication. This allows the replication of the vector to high copy numbers in cells (e.g. COS cells) which express the T antigen required for the initiation of viral DNA synthesis. Any other mammalian cell line can also be utilized for this purpose.

A typical mammalian expression vector contains the promoter element, which mediates the initiation of transcription of mRNA, the protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription can be achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses,

WO 98/44112

PCT/US98/06022

59

e.g. RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular signals can also be used (e.g. human actin promoter). Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146) and pBC12MI (ATCC 67109). Mammalian host cells that could be used include, human Hela, 283, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, African green monkey cells, quail QC1-3 cells, mouse L cells and Chinese hamster ovary cells.

Alternatively, the gene can be expressed in stable cell lines that contain the gene integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) is a useful marker to develop cell lines that carry several hundred or even several thousand copies of the gene of interest. Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy *et al.*, *Biochem J.* 227:277-279 (1991); Bebbington *et al.*, *Bio/Technology* 10:169-175 (1992)). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) cells are often used for the production of proteins.

The expression vectors pC1 and pC4 contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen *et al.*, *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart *et al.*, *Cell* 41:521-530 (1985)). Multiple cloning sites, e.g. with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors contain in addition the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene.

WO 98/44112

PCT/US98/06022

60

Example 3(a): Cloning and Expression of CAPP in COS7 Cells

The expression of plasmid, human CAPP HA is derived from a vector pCDNA3 (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) *E.coli* replication origin, 4) CMV promoter followed by a polylinker region, a SV40 intron and polyadenylation site. A DNA fragment encoding the entire human CAPP precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed by the CMV promoter. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson *et al.*, *Cell* 37:767 (1984)). The inclusion of HA tag to our target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows. The DNA sequence encoding human CAPP, ATCC # 97729, is constructed by PCR on the original clone using two primers.

Suitable primers include the following, which are used in this example.

The 5' primer has the sequence:

5' GGCC GGATCC GCC ATC ATG AGT GTT GGA CGT CGA AGA AT 3' (SEQ ID NO:6) containing the underlined BamHI restriction enzyme site followed by a Kozak sequence, followed by 23 nucleotides of coding sequence of CAPP protein in Figure 1. Inserted into an expression vector, as described below, the 5' end of the amplified fragment encoding CAPP provides an efficient signal peptide.

The 3' primer has the sequence:

5' ATTGTCTAGA ATT TTA AGC GTA GTC TGG GAC GTC GTA TGG GTA GCA TTT TAA ATG AGC ACT CTG 3' (SEQ ID NO:7) containing the underlined XbaI restriction site followed by translation stop codon, HA tag and nucleotides of CAPP sequence, including nucleotides of the coding domain set out in Figure 1.

WO 98/44112

PCT/US98/06022

61

Therefore, the PCR product contains a BamHI site, human CAPP coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and a XbaI site. The PCR amplified DNA fragment is digested with the restriction enzyme BamHI and XbaI and the vector, pcDNA3 is digested with BamHI restriction enzyme and ligated. The ligation mixture is transformed into *E. coli* strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037). The transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

For expression of the recombinant CAPP, COS cells are transfected with the expression vector by DEAE-DEXTRAN method. (J. Sambrook, E. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Laboratory Press, (1989)). Expression of the CAPP-HA protein is detected by radiolabelling and immunoprecipitation method. (E. Harlow, D. Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, (1988)). Cells are labelled for 8 hours with ³⁵S-cysteine two days post transfection. Culture media are then collected and cells are lysed with detergent (RIPA buffer (150 mM NaCl, 1% NP-40, 0.1 SDS, 1 NP-40, 0.08 DOC, 50 mM Tris, pH 7.5). (Wilson, I. *et al.*, *Cell* 37:767 (1984)). Both cell lysate and culture media are precipitated with a HA specific monoclonal antibody. Proteins precipitated are analyzed on 15 SDS-PAGE gels.

Example 3(b): Cloning and Expression in CHO Cells

The vector pC4 is used for the expression of CAPP protein. Plasmid pC4 is a derivative of the plasmid pSV2-dhfr [ATCC Accession No. 37146]. The plasmid contains the mouse DHFR gene under control of the SV40 early promoter. Chinese hamster ovary or other cells lacking dihydrofolate activity that are transfected with these plasmids can be selected by growing the cells in a selective medium (alpha minus MEM, Life Technologies) supplemented with the chemotherapeutic agent methotrexate. The amplification of the DHFR genes in cells resistant to methotrexate (MTX) has been well documented (see, e.g., Alt,

WO 98/44112

PCT/US98/06022

62

5 F.W., Kellems, R.M., Bertino, J.R., and Schimke, R.T., *J. Biol. Chem.* 253:1357-1370 (1978), Hamlin, J.L. and Ma, C., *Biochim. et Biophys. Acta*, 1097:107-143 (1990), Page, M.J. and Sydenham, M.A., *Biotechnology* 9:64-68 (1991)). Cells grown in increasing concentrations of MTX develop resistance to the drug by overproducing the target enzyme, DHFR, as a result of amplification of the DHFR gene. If a second gene is linked to the DHFR gene it is usually co-amplified and over-expressed. It is state of the art to develop cell lines carrying more than 1,000 copies of the genes. Subsequently, when the methotrexate is withdrawn, cell lines contain the amplified gene integrated into the chromosome(s).

10 Plasmid pC4 contains for the expression of the gene of interest a strong promoter of the long terminal repeat (LTR) of the Rouse Sarcoma Virus (Cullen *et al.*, *Molecular and Cellular Biology*, March 1985, 438-4470) plus a fragment isolated from the enhancer of the immediate early gene of human cytomegalovirus (CMV) (Boshart *et al.*, *Cell* 41:521-530, (1985)). Downstream of the promoter is a restriction enzyme site that allows the integration of a gene of interest. Behind the cloning site the plasmid contains translational stop codons in all three reading frames followed by the 3' intron and the polyadenylation site of the rat preproinsulin gene. Other high efficient promoters can also be used for the expression, e.g., the human β -actin promoter, the SV40 early or late promoters or the long terminal repeats from other retroviruses, e.g., HIV and HTLV. For the polyadenylation of the mRNA other signals, e.g., from the human growth hormone or globin genes can be used as well.

20 Stable cell lines carrying a gene of interest integrated into the chromosomes can also be selected upon co-transfection with a selectable marker such as gpt, G418 or hygromycin. It is advantageous to use more than one selectable marker in the beginning, e.g. G418 plus methotrexate.

30 The plasmid pC4 is digested with the restriction enzymes BamHI and XbaI and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

WO 98/44112

PCT/US98/06022

63

The DNA sequence encoding CAPP, ATCC 97729, is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene:

The 5' primer has the sequence:

5' GGCC GGATCC GCC ATC ATG AGT GTT GGA CGT CGA AGA AT 3'
(SEQ ID NO:6) containing the underlined BamHI restriction enzyme site followed by an optimized Kozak sequence for insect cell expression, followed by 23 nucleotides of coding sequence of CAPP protein in Figure 1. Inserted into an expression vector, as described below, the 5' end of the amplified fragment encoding CAPP provides an efficient signal peptide. An efficient signal for initiation of translation in eukaryotic cells, as described by Kozak, M., *J. Mol. Biol.* 196: 947-950 (1987) is appropriately located in the vector portion of the construct.

The 3' primer has the sequence:

15 5' ATTGTCTAGA ATT TTA AGC GTA GTC TGG GAC GTC GTA
TGG GTA GCA TTT TAA ATG AGC ACT CTG 3' (SEQ ID NO:7)
containing the underlined XbaI restriction site followed by translation stop codon, HA tag and nucleotides of CAPP sequence, including nucleotides of the coding domain set out in Figure 1.

20 The amplified fragments are isolated from a 1% agarose gel as described above and then digested with the endonucleases BamHI and XbaI and then purified again on a 1% agarose gel.

The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 cells are then transformed and bacteria identified that contained the plasmid pC4 inserted in the correct orientation using the restriction enzyme BamHI. The sequence of the inserted gene is confirmed by DNA sequencing.

Transfection of CHO-DHFR-cells

Chinese hamster ovary cells lacking an active DHFR enzyme are used for transfection. 5 µg of the expression plasmid C4 are cotransfected with 0.5 µg of

WO 98/44112

PCT/US98/06022

64

the plasmid pSVneo using the lipofecting method (Felgner *et al.*, supra). The plasmid pSV2-neo contains a dominant selectable marker, the gene neo from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) and cultivated from 10-14 days. After this period, single clones are trypsinized and then seeded in 6-well petri dishes using different concentrations of methotrexate (25 nM, 50 nM, 100 nM, 200 nM, 400 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (500 nM, 1 μ M, 2 μ M, 5 μ M). The same procedure is repeated until clones grow at a concentration of 100 μ M.

The expression of the desired gene product is analyzed by Western blot analysis and SDS-PAGE.

Example 4: Tissue distribution of CAPP protein expression

Northern blot analysis is carried out to examine CAPP gene expression in human tissues, using methods described by, among others, Sambrook *et al.*, cited above. A cDNA probe containing the entire nucleotide sequence of the CAPP protein (SEQ ID NO:1) is labeled with 32 P using the *rediprime*TM DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labelling, the probe is purified using a CHROMA SPIN-100TM column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labelled probe is then used to examine various human tissues for CAPP mRNA.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) are obtained from Clontech and are examined with labelled probe using ExpressHybTM hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C.

WO 98/44112

PCT/US98/06012

65

5 overnight, and films developed according to standard procedures. By Northern blot analysis it has been determined that this gene is abundant in adult heart and pancreas, with low amounts in placenta, lung, liver, skeletal muscle, kidney, spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocytes. The gene was identified by database distribution in activated T cells (3), CD34 positive cells, Ntera2 cells 14 days after RA stimulation, kidney cortex, adult heart, Jurkat cells and small intestine.

10 It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

15 The entire disclosure of all publications (including patents, patent applications, journal articles, laboratory manuals, books, or other documents) cited herein are hereby incorporated by reference.

WO 98/44112

PCT/US98/06022

-66-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HUMAN GENOME SCIENCES, INC.
9410 KEY WEST AVENUE
ROCKVILLE, MD 20850
UNITED STATES OF AMERICA

APPLICANT/INVENTOR: SOPPET, DANIEL R.
RUBEN, STEVEN M.

(ii) TITLE OF INVENTION: CARDIAC AND PANCREATIC PROTEIN AND GENE

(iii) NUMBER OF SEQUENCES: 42

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
(B) STREET: 1100 NEW YORK AVENUE, SUITE 600
(C) CITY: WASHINGTON
(D) STATE: DC
(E) COUNTRY: US
(F) ZIP: 20005-3934

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: To be assigned
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/042,855
(B) FILING DATE: 28-MAR-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: STEFFE, ERIC K.
(B) REGISTRATION NUMBER: 36,688
(C) REFERENCE/DOCKET NUMBER: 1488.062PC01

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (202) 371-2600
(B) TELEFAX: (202) 371-2540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2745 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

WO 98/44112

PCT/US98/06022

-67-

(A) NAME/KEY: CDS
(B) LOCATION: 233..1423

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 233..328

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION: 329..1423

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

GCAGCGGCAC GGCAGCAGCG GCAACAAGTG CCGGACTAGC AGAGCCAAGC CGGAGCAGTC      60
CCTGCCGCCG ACACCGCCCG GCGGCCCGTC CGGGCGCCCG CGCATGGAGC GTGAGCTGCG      120
GCGGTCGCCG GGGCTGAGCC GCGCGGAGCG CCGGGACGTG GATGTGGCCG CGATCTCCCG      180
CCCTTGCCCC CGCCCCGCCG AGCTGGAGCT GCTCCCGGAC AAGATATGAG AA ATG      235
                               Met
                               -32

AGT GTT GGA CGT CGA AGA ATA AAG TTG TTG GGT ATC CTG ATG ATG GCA      283
Ser Val Gly Arg Arg Arg Ile Lys Leu Leu Gly Ile Leu Met Met Ala
-30                               -25                               -20

AAT GTC TTC ATT TAT TTT ATT ATG GAA GTC TCC AAA AGC AGT AGC CAA      331
Asn Val Phe Ile Tyr Phe Ile Met Glu Val Ser Lys Ser Ser Ser Gln
-15                               -10                               -5                               1

GAA AAA AAT GGA AAA GGG GAA GTA ATA ATA CCC AAA GAG AAG TTC TGG      379
Glu Lys Asn Gly Lys Gly Glu Val Ile Ile Pro Lys Glu Lys Phe Trp
                               5                               10                               15

AAG ATA TCT ACC CCT CCC GAG GCA TAC TGG AAC CGA GAG CAA GAG AAG      427
Lys Ile Ser Thr Pro Pro Glu Ala Tyr Trp Asn Arg Glu Gln Glu Lys
                               20                               25                               30

CTG AAC CGG CAG TAC AAC CCC ATC CTG AGC ATG CTG ACC AAC CAG ACG      475
Leu Asn Arg Gln Tyr Asn Pro Ile Leu Ser Met Leu Thr Asn Gln Thr
                               35                               40                               45

GGG GAG GCG GGC AGG CTC TCC AAT ATA AGC CAT CTG AAC TAC TGC GAA      523
Gly Glu Ala Gly Arg Leu Ser Asn Ile Ser His Leu Asn Tyr Cys Glu
50                               55                               60                               65

CCT GAC CTG AGG GTC ACG TCG GTG GTT ACG GGT TTT AAC AAC TTG CCG      571
Pro Asp Leu Arg Val Thr Ser Val Val Thr Gly Phe Asn Asn Leu Pro
                               70                               75                               80

GAC AGA TTT AAA GAC TTT CTG CTG TAT TTG AGA TGC CGC AAT TAT TCA      619
Asp Arg Phe Lys Asp Phe Leu Leu Tyr Leu Arg Cys Arg Asn Tyr Ser
                               85                               90                               95

CTG CTT ATA GAT CAG CCG GAT AAG TGT GCA AAG AAA CCT TTC TTG TTG      667
Leu Leu Ile Asp Gln Pro Asp Lys Cys Ala Lys Lys Pro Phe Leu Leu
100                               105                               110

CTG GCG ATT AAG TCC CTC ACT CCA CAT TTT GCC AGA AGG CAA GCA ATC      715
Leu Ala Ile Lys Ser Leu Thr Pro His Phe Ala Arg Arg Gln Ala Ile

```

SDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

-68-

115	120	125	
CGG GAA TCC TGG GGC CAA GAA AGC AAC GCA GGG AAC CAA ACG GTG GTG			763
Arg Glu Ser Trp Gly Gln Glu Ser Asn Ala Gly Asn Gln Thr Val Val			
130	135	140	145
CGA GTC TTC CTG CTG GGC CAG ACA CCC CCA GAG GAC AAC CAC CCC GAC			811
Arg Val Phe Leu Leu Gly Gln Thr Pro Pro Glu Asp Asn His Pro Asp			
150	155	160	
CTT TCA GAT ATG CTG AAA TTT GAG AGT GAG AAG CAC CAA GAC ATT CTT			859
Leu Ser Asp Met Leu Lys Phe Glu Ser Glu Lys His Gln Asp Ile Leu			
165	170	175	
ATG TGG AAC TAC AGA GAC ACT TTC TTC AAC TTG TCT CTG AAG GAA GTG			907
Met Trp Asn Tyr Arg Asp Thr Phe Phe Asn Leu Ser Leu Lys Glu Val			
180	185	190	
CTG TTT CTC AGG TGG GTA AGT ACT TCC TGC CCA GAC ACT GAG TTT GTT			955
Leu Phe Leu Arg Trp Val Ser Thr Ser Cys Pro Asp Thr Glu Phe Val			
195	200	205	
TTC AAG GGC GAT GAC GAT GTT TTT GTG AAC ACC CAT CAC ATC CTG AAT			1003
Phe Lys Gly Asp Asp Asp Val Phe Val Asn Thr His His Ile Leu Asn			
210	215	220	225
TAC TTG AAT AGT TTA TCC AAG ACC AAA GCC AAA GAT CTC TTC ATA GGT			1051
Tyr Leu Asn Ser Leu Ser Lys Thr Lys Ala Lys Asp Leu Phe Ile Gly			
230	235	240	
GAT GTG ATC CAC AAT GCT GGA CCT CAT CGG GAT AAG AAG CTG AAG TAC			1099
Asp Val Ile His Asn Ala Gly Pro His Arg Asp Lys Lys Leu Lys Tyr			
245	250	255	
TAC ATC CCA GAA GTT GTT TAC TCT GGC CTC TAC CCA CCC TAT GCA GGG			1147
Tyr Ile Pro Glu Val Val Tyr Ser Gly Leu Tyr Pro Tyr Ala Gly			
260	265	270	
GGA GGG GGG TTC CTC TAC TCC GGC CAC CTG GCC CTG AGG CTG TAC CAT			1195
Gly Gly Gly Phe Leu Tyr Ser Gly His Leu Ala Leu Arg Leu Tyr His			
275	280	285	
ATC ACT GAC CAG GTC CAT CTC TAC CCC ATT GAT GAC GTT TAT ACT GGA			1243
Ile Thr Asp Gln Val His Leu Tyr Pro Ile Asp Asp Val Tyr Thr Gly			
290	295	300	305
ATG TGC CTT CAG AAA CTC GGC CTC GTT CCA GAG AAA CAC AAA GGC TTC			1291
Met Cys Leu Gln Lys Leu Gly Leu Val Pro Glu Lys His Lys Gly Phe			
310	315	320	
AGG ACA TTT GAT ATC GAG GAG AAA AAC AAA AAT AAC ATC TGC TCC TAT			1339
Arg Thr Phe Asp Ile Glu Glu Lys Asn Lys Asn Ile Cys Ser Tyr			
325	330	335	
GTA GAT CTG ATG TTA GTA CAT AGT AGA AAA CCT CAA GAG ATG ATT GAT			1387
Val Asp Leu Met Leu Val His Ser Arg Lys Pro Gln Glu Met Ile Asp			
340	345	350	
ATT TGG TCT CAG TTG CAG AGT GCT CAT TTA AAA TGC TAAAATAGAT			1433
Ile Trp Ser Gln Leu Gln Ser Ala His Leu Lys Cys			
355	360	365	

VSDOCID: <WO 9844112A1>

WO 98/44112

PCT/US98/06022

-69-

ACAAACTCAA TTTGTCATAG AAAGGTGTAT TTTGAATAGT TCCCATGTTG TGTCTCACA 1493
 TTAGAGTAAT TTCTATATTA AACCATGAAA ATTGCCTTTA TGAGTGATAC CCATTTGAGG 1553
 GCCTCTAAAC CCTTCAATTT GCTACTCAGC TGAAGAGGGA AAGCGGAAGA TGSTAATTTT 1613
 TTTTATGGA TGATATGGCA GGATGATTGG TTCTGATCTT ACCGGCTAGT GGTCAATTTT 1673
 AAAAACTTG TACCCTCTTA TCTGAAATCC TGTTCCTGGA ATTTGGCCAT TTTAAGTGAT 1733
 TTTGTTTGCC CTCTCTATA ATATTCCTAC TTCCCATAAAT AATGACTGAT TTATTGTAA 1793
 TTCAGGTATT TATAAACCTA TTGGCTACAA AGACTTTGTT AAACATTATC CAGTGGTTTT 1853
 CGTGAAATGG AATTATGTTT ATTTTATGGA GATTGCGTA AATTTTAAAT TGTCTAGAAA 1913
 ACTGAAATTT CAGTTGTCAG TTGTGGAATT CAGTTTTTCA ATTGTGGAAA TTTCTGCCA 1973
 CCCCAACAGT ATTTTGTGT GTTAATTAAT TTGCAAAAT GAGAATCATG GTGTGACACT 2033
 CATCTAATTT ATCTTGTGT GATGTTATGG TCATAATAAG GAGAAAGAGG GTTTAATTTT 2093
 TCTGTATTT GGTTCCTGG TGGTATCATA GTGTAATTTT AGTATTTGAA AATCAGTGTG 2153
 ATTCCTTAAT GGCCAACTGA AGATTGAATT GCCGCTAACA ACCATATCGT GTTAGTGAAT 2213
 TTTCAATATG GACCAGGAAG GCATATGTAT TTTGAACTTG AGTGAAAAGG TTGAAGTTAC 2273
 AGACTTTTGC ATAGATGGTT TGTCAATTTA AAATTCAGA ATTTATTATT GCCATATTTT 2333
 CACATGCTGC TTATACAAGA TTATTATTGA GTAGTAACTG TTCCCTGTCT ATGTAGAAGT 2393
 GCCTGTGTTT TTATTTATTG TTCCAGATCA AAGACCAAAA CATTTTCTTA AATATCTCTT 2453
 ATGTAATATT TTATTTGTAT ACAGTGTGT TGATGAAATA TTTAACTAGA GCATGATATT 2513
 TTAAATGTTA AGGTGTAACA TAIGTTAAAT AAACTGTTA TTTTGAATT TTAATAATTG 2573
 TTTTTGGGG GTATGAACTA CTAGAGTTTA AAATCTGCC AACTATTAC TTATATGTAC 2633
 TATTGTGTA CATACTTTCT TGAAATATTT TTGTTATAG AATTGAAGGT TCTTATCAGA 2693
 TGGGATACTG GGGATTATAA ACAATGGAAA TAAAGCCACT GTATTTTAA AA 2745

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 397 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ser Val Gly Arg Arg Arg Ile Lys Leu Leu Gly Ile Leu Met Met
 -32 -30 -25 -20
 Ala Asn Val Phe Ile Tyr Phe Ile Met Glu Val Ser Lys Ser Ser Ser
 -15 -10 -5

PCT/US98/06022

WO 98/44112

-70-

Gln Glu Lys Asn Gly Lys Gly Glu Val Ile Ile Pro Lys Glu Lys Phe
 1 5 10 15
 Trp Lys Ile Ser Thr Pro Pro Glu Ala Tyr Trp Asn Arg Glu Gln Glu
 20 25 30
 Lys Leu Asn Arg Gln Tyr Asn Pro Ile Leu Ser Met Leu Thr Asn Gln
 35 40 45
 Thr Gly Glu Ala Gly Arg Leu Ser Asn Ile Ser His Leu Asn Tyr Cys
 50 55 60
 Glu Pro Asp Leu Arg Val Thr Ser Val Val Thr Gly Phe Asn Asn Leu
 65 70 75 80
 Pro Asp Arg Phe Lys Asp Phe Leu Leu Tyr Leu Arg Cys Arg Asn Tyr
 85 90 95
 Ser Leu Leu Ile Asp Gln Pro Asp Lys Cys Ala Lys Lys Pro Phe Leu
 100 105 110
 Leu Leu Ala Ile Lys Ser Leu Thr Pro His Phe Ala Arg Arg Gln Ala
 115 120 125
 Ile Arg Glu Ser Trp Gly Gln Glu Ser Asn Ala Gly Asn Gln Thr Val
 130 135 140
 Val Arg Val Phe Leu Leu Gly Gln Thr Pro Pro Glu Asp Asn His Pro
 145 150 155 160
 Asp Leu Ser Asp Met Leu Lys Phe Glu Ser Glu Lys His Gln Asp Ile
 165 170 175
 Leu Met Trp Asn Tyr Arg Asp Thr Phe Phe Asn Leu Ser Leu Lys Glu
 180 185 190
 Val Leu Phe Leu Arg Trp Val Ser Thr Ser Cys Pro Asp Thr Glu Phe
 195 200 205
 Val Phe Lys Gly Asp Asp Asp Val Phe Val Asn Thr His His Ile Leu
 210 215 220
 Asn Tyr Leu Asn Ser Leu Ser Lys Thr Lys Ala Lys Asp Leu Phe Ile
 225 230 235 240
 Gly Asp Val Ile His Asn Ala Gly Pro His Arg Asp Lys Lys Leu Lys
 245 250 255
 Tyr Tyr Ile Pro Glu Val Val Tyr Ser Gly Leu Tyr Pro Pro Tyr Ala
 260 265 270
 Gly Gly Gly Gly Phe Leu Tyr Ser Gly His Leu Ala Leu Arg Leu Tyr
 275 280 285
 His Ile Thr Asp Gln Val His Leu Tyr Pro Ile Asp Asp Val Tyr Thr
 290 295 300
 Gly Met Cys Leu Gln Lys Leu Gly Leu Val Pro Glu Lys His Lys Gly
 305 310 315 320
 Phe Arg Thr Phe Asp Ile Glu Glu Lys Asn Lys Asn Asn Ile Cys Ser
 325 330 335

WO 98/44112

PCT/US98/06022

-71-

Tyr Val Asp Leu Met Leu Val His Ser Arg Lys Pro Gln Glu Met Ile
 340 345 350

Asp Ile Trp Ser Gln Leu Gln Ser Ala His Leu Lys Cys
 355 360 365

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 323 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gln Ser Lys His Arg Lys Leu Leu Leu Arg Cys Leu Leu Val Leu Pro
 1 5 10 15
 Leu Ile Leu Leu Val Asp Tyr Cys Gly Leu Leu Thr His Leu His Glu
 20 25 30
 Leu Asn Phe Glu Arg His Phe His Tyr Pro Leu Asn Asp Asp Thr Gly
 35 40 45
 Ser Gly Ser Ala Ser Ser Gly Leu Asp Lys Phe Ala Tyr Leu Arg Val
 50 55 60
 Pro Ser Phe Thr Ala Glu Val Pro Val Asp Gln Pro Ala Arg Leu Thr
 65 70 75 80
 Met Leu Ile Lys Ser Ala Val Gly Asn Ser Arg Arg Arg Glu Ala Ile
 85 90 95
 Arg Arg Thr Trp Gly Tyr Glu Gly Arg Phe Ser Asp Val His Leu Arg
 100 105 110
 Arg Val Phe Leu Leu Gly Thr Ala Glu Asp Ser Glu Lys Asp Val Ala
 115 120 125
 Trp Glu Ser Arg Glu His Gly Asp Ile Leu Gln Ala Asp Phe Thr Asp
 130 135 140
 Ala Tyr Phe Asn Asn Thr Leu Lys Thr Met Leu Gly Met Arg Trp Ala
 145 150 155 160
 Ser Glu Gln Phe Asn Arg Ser Glu Phe Tyr Leu Phe Val Asp Asp Asp
 165 170 175
 Tyr Tyr Val Ser Ala Lys Asn Val Leu Lys Phe Leu Gly Arg Gly Arg
 180 185 190
 Gln Ser His Gln Pro Glu Leu Leu Phe Ala Gly His Val Phe Gln Thr
 195 200 205
 Ser Pro Leu Arg His Lys Phe Ser Lys Trp Tyr Val Ser Leu Glu Glu
 210 215 220

PCT/US98/06022

WO 98/44112

-72-

Tyr Pro Phe Asp Arg Trp Pro Pro Tyr Val Thr Ala Gly Ala Phe Ile
 225 230 235 240
 Leu Ser Gln Lys Ala Leu Arg Gln Leu Tyr Ala Ala Ser Val His Leu
 245 250 255
 Pro Leu Phe Arg Phe Asp Asp Val Tyr Leu Gly Ile Val Ala Leu Lys
 260 265 270
 Ala Gly Ile Ser Leu Gln His Cys Asp Asp Phe Arg Phe His Arg Pro
 275 280 285
 Ala Tyr Lys Gly Pro Asp Ser Tyr Ser Ser Val Ile Ala Ser His Glu
 290 295 300
 Phe Gly Asp Pro Glu Glu Met Thr Arg Val Trp Asn Glu Cys Arg Ser
 305 310 315 320
 Ala Asn Tyr

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AGCAGGATCC CAAGAAAAAA ATGGAAAAGG G

31

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTGTCTAGA TATCTATTTT AGCATTTTA

29

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 39 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

WO 98/44112

PCT/US98/06022

-73-

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGCCGGATCC GCCATCATGA GTGTTGGACG TCGAAGAAT

39

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 64 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATTGTCTAGA ATTTTAAGCG TAGTCTGGGA CGTCGTATGG GTAGCATTTT AAATGAGCAC
TCTG

60

64

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3974 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGTACCTAAG TGAGTAGGGC GTCCGATCGA CGGACGCCCTT TTTTTTGAAT TCGTAATCAT 60
GGTCATAGCT GTTTCCTGTG TGAAATTGTT ATCCGCTCAC AATCCACAC AACATACGAG 120
CCGGAAGCAT AAAGTGTAAG GCCTGGGGTG CCTAATGAGT GAGCTAACTC ACATTAATTG 180
CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG GAAACCTGTC GTGCCAGCTG CATTAAATGAA 240
TCGGCCAACG GCGGGGGAGA GGCGGTTTGC GTATTGGGCG CTCTCCGCT TCCTCGCTCA 300
CTGACTCGCT GCGCTCGGTC GTTCGGCTGC GCGGAGCGGT ATCAGCTCAC TCAAAGGCGG 360
TAATACGGTT ATCCACAGAA TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAGGCC 420
AGCAAAGGC CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTCCAT AGGCTCCGCC 480
CCCCTGACGA GCATCACAAG AATCGACGCT CAAGTCAGAG GTGGCGAAAC CCGACAGGAC 540
TATAAGATA CCAGGCGTTT CCCCTGGAA GCTCCCTCGT GCGCTCTCCT GTCCGACCC 600

WO 98/44112

PCT/US98/06022

-74-

TGCCGCTTAC CGGATACCTG TCCGCCTTTC TCCCTTCGGG AAGCGTGGCG CTTTCTCATA	660
GCTCACGCTG TAGGTATCTC AGTTCGGTGT AGGTCGTTCC CTCCAAGCTG GGCTGTGTGC	720
ACGAACCCCC CGTTCAGCCC GACCGCTGCG CTTATCCGG TAACTATCGT CTTGAGTCCA	780
ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG ATTAGCAGAG	840
CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG GCCTAACTAC GGCTACACTA	900
GAAGAACAGT ATTTGGTATC TCGGCTCTGC TGAAGCCAGT TACCTTCGGA AAAAGAGTTG	960
GTAGCTCTTG ATCCGGCAAA CAAACCACCG CTGGTAGCGG TGTTTTTTTT GTTTGCAAGC	1020
AGCAGATTAC GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT	1080
CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCAAGAGA TTATCGTCGA	1140
CAATTCGCGC GCGAAGGCGA AGCGGCATGC ATTTACGTTG ACACCATCGA ATGGTGCAAA	1200
ACCTTTCGCG GTATGGCATG ATAGCGCCCG GAAGAGAGTC AATTCAGGGT GGTGAATGTG	1260
AAACCAGTAA CGTTATACGA TGTCGCAGAG TATGCCGGTG TCTCTTATCA GACCGTTTCC	1320
CGCGTGCTGA ACCAGGCCAG CCACGTTTCT GCGAAAACGC GGGAAAAAGT GGAAGCGGCG	1380
ATGGCGGAGC TGAATTACAT TCCCAACCGC GTGGCACAAC AACTGGCGGG CAAACAGTCG	1440
TTGCTGATTG GCGTTGCCAC CTCCAGTCTG GCCCTGCACG CGCCGTCGCA AATTGTCGCG	1500
GCGATTAAAT CTCGGCCGA TCAACTGGGT GCCAGCGTGG TGGTGTCGAT GGTAGAACGA	1560
AGCGGCGTCG AAGCCTGTAA AGCGGCGGTG CACAATCTTC TCGCGCAACG CGTCAGTGGG	1620
CTGATCATTG ACTATCCGCT GGATGACCAG GATGOCATTG CTGTGGAAGC TGCCTGCACT	1680
AATGTTCCGG CGTTATTTCT TGATGTCTCT GACCAGACAC CCATCAACAG TATTATTTTC	1740
TCCCATGAAG ACGGTACGCG ACTGGGCGTG GAGCATCTGG TCGCATTGGG TCACCAGCAA	1800
ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGGCGC GTCTGCGTCT GGCTGGCTGG	1860
CATAAATATC TCACTCGCAA TCAAATTCAG CCGATAGCGG AACGGGAAGG CGACTGGAGT	1920
GCCATGTCCG GTTTTCAACA AACCATGCAA ATGCTGAATG AGGGCATCGT TCCCACTGCC	1980
ATGCTGGTTG CCAACGATCA GATGGCGCTG GCGCAATGC GCGCCATTAC CGAGTCCGGG	2040
CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACG ATACCGAAGA CAGCTCATGT	2100
TATATCCCGC CGTTAACCAC CATCAAACAG GATTTTCGCC TGCTGGGGCA AACCAGCGTG	2160
GACCGCTTGC TGCAACTCTC TCAGGGCCAG GCGGTGAAGG GCAATCAGCT GTTGCCCGTC	2220
TCACTGGTGA AAAGAAAAAC CACCCTGGCG CCCAATACGC AAACCGCCTC TCCCGCGCG	2280
TTGCGCGATT CATTAATGCA GCTGGCAGCA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA	2340
GCGCAACGCA ATTAATGTAA GTTAGCGCGA ATTGTCGACC AAAGCGGCCA TCGTGCCCTC	2400
CCACTCCTGC AGTTCGGGGG CATGGATGCG CGGATAGCGG CTGCTGGTTT CCTGGATGCC	2460

ISDOCD: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

-75-

GACGGATTTG CACTGCCGGT AGAACTCCGC GAGGTCGTCC AGCCTCAGGC AGCAGCTGAA 2520
CCAACTCGCG AGGGGATCGA GCCCGGGGTG GCGGAAGAAC TCCAGCATGA GATCCCCGCG 2580
CTGGAGGATC ATCCAGCCGG CGTCCCGGAA AACGATTCCG AAGCCCAACC TTTCATAGAA 2640
GGCGGGGGTG GAATCGAAAT CTCGTGATGG CAGGTTGGGC GTCGCTTGGT CGGTCAATTC 2700
GAACCCAGAG GTCCCGCTCA GAAGAACTCG TCAAGAAGGC GATAGAAGGC GATGCGCTGC 2760
GAATCGGGAG CGGCGATACC GTAAAGCAG AGGAAGCGGT CAGCCCATTC GCCGCCAAGC 2820
TCTTCAGCAA TATCACGGGT AGCCAACGCT ATGTCCTGAT AGCGGTCCGC CACACCCAGC 2880
CGGCCACAGT CGATGAATCC AGAAAAGCGG CCATTTTCCA CCATGATATT CGGCAAGCAG 2940
GCATCGCCAT GGGTCACGAC GAGATCCTCG CCGTCGGGCA TCGCGCCCTT GAGCCTGGCG 3000
AACAGTTCGG CTGGCGCGAG CCCCTGATGC TCTTCGTCCA GATCATCTG ATCGACAAGA 3060
CCGCTTCCA TCCGAGTACG TGCTCGCTCG ATGCGATGTT TCGCTTGGTG GTCGAATGGG 3120
CAGGTAGCCG GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT GGATACTTTC 3180
TCGGCAGGAG CAAGGTGAGA TGACAGGAGA TCCTGCCCCG GCACTTCGCC CAATAGCAGC 3240
CAGTCCCTTC CCGCTTCAGT GACAACTCG AGCACAGCTG CGCAAGGAAC GCCCGTCGTG 3300
GCCAGCCACG ATAGCCGCGC TGCCTCGTCC TGCACTTCAT TCAGGGCACC GGACAGGTGC 3360
GTCTTGACAA AAAGAACCGG GCGCCCTGC GCTGACAGCC GGAACACGGC GGCATCAGAG 3420
CAGCCGATTG TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA AGCGGCCGGA 3480
GAACCTGCGT GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC TGTCTCTGA 3540
TCAGATCTTG ATCCCTGCG CCATCAGATC CTTGGCGGCA AGAAAGCCAT CCAGTTTACT 3600
TTGCAGGGCT TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCGG TCGCTTGCT 3660
GTCCATAAAA CCGCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC TACCTGCTTT 3720
CTCTTTGCGC TTGCGTTTC CTTGTCCAG ATAGCCAGT AGCTGACATT CATCCGGGT 3780
CAGCACCGTT TCTGCGGACT GGCTTTCTAC GTGTTCCGCT TCCTTTAGCA GCCCTTGCGC 3840
CCTGAGTGCT TGCGGCAGCG TGAAGCTTAA AAACTGCAA AAAATAGTTT GACTTGTGAG 3900
CGGATAACAA TTAAGATGTA CCCAATTGTG AGCGGATAAC AATTTACAC ATTAAGAGG 3960
AGAAATTACA TATG 3974

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 112 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

IDOCID: <WO_984112A1J>

WO 98/44112

PCT/US98/06022

-76-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AAGCTTAAAA AACTGCAAAA AATAGTTTGA CTTGTGAGCG GATAACAATT AAGATGTACC 60
CAATTGTGAG CGGATAACAA TTTCACACAT TAAAGAGGAG AAATTACATA TG 112

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 299 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AAAAATACAG TGGCTTTATT TCCATTGTTT ATAGTCCCCA GTATCCCATC TGATAAGAAC 60
CTTCAATTCT ATAAACAAAA ATATTTCAAG AAAGTATGTT ACACAATAGT ACATATAAGT 120
AATAGTTTGG CAGAATTTTA AACTCTAGTA GTTCATACCC CCAAAAAACA AATTTTAAAN 180
TTCAAAAATA ACAGTTTAT TTAACATATG TTACACCTTA ACATTTAAAA TATCATGCTC 240
TAGTTAAATA TTTTCATCAAC AACACTGTAT ACANNTAAAA TATTACATAA AATATATTT 299

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTTGCCAGA AGGCAAGCAA TCCGGGAATC CTGGGGCCAA GAAAGCAACG CAGGGAACCA 60
AACGGTGCTG CGAGTNTTCC TGCTGGGCCA GACACCCCA GAGGACAACC ACCCCGACCT 120
TTCAGATATG CTGAAATTTG AGAGTGAGAA GCACCAAGAC ATTCTTATGT GGAACACAG 180
AGACACTTTN TTCAACTTGT CTCTGAAGGA AGTGCTGTTT CTNAGGTGGG TAAGTACTTC 240
CTGCCCAGAC ACTGAGTTTG TTTTCAAGGG CGATGACGAT GT 282

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 266 base pairs

WO 98/44112

PCT/US98/06022

-77-

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TTTGGCCAGA AGGCAAGCAA TCCGGGAATC CTGGGGCCAA GAAAGCAACG CAGGGAACCA	60
AACGGTGGTG CGAGTNTTCC TGCTGGGCCA GACACCCCCA GAGGACAACC ACCCCGACCT	120
TTCAGATATG CTGAAATTTG AGAGTNAGAA GCACCAAGAC ATTCTTATGT GGAAGTACAG	180
AGACACTTTC TTCAACTTGT CTCTGAAGGA AGTGCTGTTT CTCAGGTGGG TAAGTACTTC	240
CTGCCCAGAC ACTGAGTTTG TTTCA	266

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTTGCAAAAT GAGAATCATG GTGTGACACT CATCTAATTT ATCTTGTTGT GATGTTATGG	60
TCATAATAAG GAGAAANAGG GTTAAATTTT NCTTGATTTT GGTTCCTGG TGGTATCATA	120
GTGTAATTTT AGTATTTGAA AATCAGTGTG ATTCCTTAAT GGCCAACTGA AGATTGAATT	180
GCCGCTAACA ACCATATCGT GTTAGTGAAT TTNCAATATG GACCAGGAAG GCATATGTAT	240
TTTGAAGTGG GAGTGAAAAG GTTGAAGTT ACAGACTTTT TGGCATAGGT GGGTTTGGTC	300
CAATTTTAAA ATTCCCGAAT TTATTNNTTG NCNNTTNTN CACATGGGNG GTTATTACAG	360
G	361

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 259 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

WO 98/44112

PCT/US98/06022

-78-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TTGTTGGGTA TCCTGATGAT GGCAAATGTC TTCATTTATT TTATTATGGA AGTCTCCAAA	60
AGCAGTAGCC AAGAAAAAAA TGGAAAAGGG GAAGTAATAA TACCCAAAGA GAAGTTCTGG	120
AAGATATCTA CCCCTCCCGA GGCATACTNG AACCGAGAGC AAGAGAAGCT GAACCGGCAG	180
TACAACCCCA TCCTGAGCAT GCTGACCAAC CAGACGGGGG AGGCGGGCAG GCTCTCCAAT	240
ATAAGNCATC TGAACACT	259

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 195 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAAAATACAG TGGCTTTATT TCCATTGTTT ATAGTCCCA GTATCCCATC TGATAAGAAC	60
CTNCAATTCT ATAAACAAA ATATTTCAAG AAAGTATGTT ACACAATAGT ACATATAAGT	120
AATAGTTTGG CAGAATTTA AACTCTAGTA GTTCATACCC CCAAAAAACA AATTTTAAAA	180
TTCAAAAATA ACAGT	195

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 521 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTATAATATT CCTACTTCCC ATAATAATGA CTGATTTATT TGTAATTCAG GTATTTATAA	60
ACCTATTGGC TACAAAGACT TTGTAAACA TTATCCAGTG GTTTTCGTGA AATGGAATTA	120
TGTTTATTTT TATGGGATTI GGGTAAATTT TAAATTGTCT AGAAACTGA AATTCAGTT	180
GTCAGTTGTG GAATTCAGTT TTTCAATTGT GGAAATTTCC TGCCACCCCA ACAGTATTTT	240
TGTGTGTTAA TTAATTTTGC AAAATGAGAA TCATGGTGTG ACACTCATCT AATTTATCTT	300
GTTGTGATGT TATGGTCATA ATACGGAGAA AGAGGGTTTA ATTTTCTTG TATTTGGTTT	360

WO 98/44112

PCT/US98/06022

-79-

CCTGGTGGTA TCATAGTGTA ATTTTAGTAT TTGAAAATCA GTGTGATTCC TTAATGGCCA 420
ACTGAAGATT GAATTGCCGC TAACAACCAT ATCGTGTTAG TGAATTTTCA ATATGGACCA 480
GGAAGGCATA TGTAATTTGA ACTTGAGTGA AAAGGTTGAA G 521

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 517 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AATAGATACA AACTCAATTT TGCATAGAAA GGTATATTTT GAATAGTTCC CATGTTGTGT 60
TCTCACATTA GAGTAATTC TGTATTAAAC CATGAAAATT GCACTTTATG ACTGATACCC 120
ATTTGAGGGC CTCTAAACCC TTCAATTGG TACTCACGTG AAGAGGGAAA GCGGAAGATG 180
GTAATTTTTT TTTACGGATG ATATGCCAGG ATGATTGGTT CTGATCTTAC CGGCTAGTGG 240
TCATTTTTTAA AAAACTTGTA CCCTCTTATC TGAAATCCTG TTTCTGGAAT TTGGCCATTT 300
TAAGTGATTT TGTTTGCCCT CTTCTATAAT ATTCTACTT CCCATAATAA TGA CTGATTT 360
ATTTGTAATT CAGGTATTTA TAAACCTATT GGCTACAAAG ACTTTGTTAA ACATTATCCA 420
GTGGTTTTTCG TGAAATGGAA TTATGTATAT TTTTATGGGA TTTGGGAAAT TTAAATTGT 480
CTAGAAAAC TGAATTTTCA TGTTCAGTTG TGAAT 517

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 462 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATACTGTTGG GGTGGCAGGA AATTTCCACA ATTGAAAAC TGAATTCAC AACTGACAAC 60
TGAAATTTCA GTTTTCTAGA CAATTTAAAA TTTACCCAAA TCCCATAAAA ATAAACATAA 120
TTCCATTTCA CGAAACCAC TGGATAATGT TTAACAAAGT CTTGTAGCC AATAGGTTTA 180
TAAATACCTG AATTACAAAT AAATCAGTCA TTATTATGGG AAGTAGGAAT ATTATAGAAG 240

WO 98/44112

PCT/US98/06022

-80-

AGGGCAAACA AAATCACTTA AAATGGCCAA ATTCCAGAAA CAGGATTTC AATAAGAGGG 300
TACAAGTTTT TTAATAATNG ACCACTAGCC GGTAAAGATCA GATCCAATCA TCCTGCCATA 360
TCATCCGTAA AAAAAAATTA CCATCTTCCG CTTTCCCTCT TCACGTGAGT ACCAAATTGG 420
AAGGGGTTAG AGGCCCTCAA ACGGGTATCA CTCATAAAGG CA 462

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CTNTAATTAA CACACAAAAA TACTGTTGGG GTGGCAGGAA ATTTCCACAA TTGAAAAACT 60
GAATTCACCA ACTGACAAC TAAATTTTCTAGAC AATTTAAAT TTACCCAAAT 120
CCCATAAAAA TAAACATAAT TCCATTTTAC GAAACCACT GGATAATGTT TAACAAAGTC 180
TTTGTAGCCA ATAGGTTTAT AAATACCTGA ATTACAAATA AATCAGTCAT TATTATGGGA 240
AGTAGGAATA TTATNGAAGA GGGCAAACAA AATCACTTAA AATGGCCAAA TTCCAGAAAC 300
AGGATTTTCTAG ATAAGAGGGT ACAAGTTTTT TAAAAATGAC CACTAGCCGG TAAGATCAGA 360
ACCAATCATC CTGCCATATC ATCCGTAAAA NAAATTACC ATCTTCCGCT TTCCCTCTTC 420
ACGTGAGTAC CAAATTGAAG GGTTTAGG 448

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 857 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

NGNGGTNNCG TCGGTTAAAT ATTCAAGACC AAAGCCAAAG ATCTCTTCAT AGGTGATGTG 60
ATCCACAATG CTGGACCTCA TCGGGATAAG AAGCTGAAGT ACTACATCCC AGAAGTTGTT 120
TACTCTGGCC TCTACCCACC CTATGCAGGG GGAGGGGGGT TCCTCTACTC CGGCCACCTG 180
GCCCTGAGGC TGTACCATAT CACTGACCAG GTCCATCTCT ACCCCATTGA TGACGTTTAT 240

WO 98/44112

PCT/US98/06022

-81-

ACTGGAATGT GCCTTCAGAA ACTCGGCCTC GTTCCAGAGA AACACAAAGG CTTCAGGACA 300
TTTGATATCG AGGAGAAAAA CAAAAATAAC ATCTGCTCCT ATGTAGATCT GATGTTAGGA 360
CATAGNAGGA AAACCTCAAG AGATGATTGA TATTTGGGCT CAAGNTGCAG AGTGCTCAAT 420
TTAAATGCT AAAATAGATA CAACTCAAT TTGGGATTNG AAGGGGTTTT TNGGATTGGC 480
CCCATNTGGG GTCTTNANAT TAGAGMNGGT TCAAGTGGCT ACAGTGATGA AAANNNNNNN 540
NNNNGGNNNN NNNNCCNNNT NNTTNMAANN NNNNNNNNNN NNNNNNNNNN NNTNNCANN 600
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN 660
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN 720
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN 780
NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNN 840
NNNNNNNNNN NNNNNNG 857

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 467 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAGATGATTG ATATTGGTC TCAGTTCAG AGTGCTCATT TAAATGCTA AAATAGATAC 60
AAACTCAATT TTGCATAGAA AGGTGTATTT TGAATAGTTC CCATGTTGTG TTCTCACATT 120
AGAGTAATTT CTGTATTAAA CCATGAAAAT TGCCTTTATG AGTGATACCC ATTTGAGGGC 180
CTCTAAACCC TTCAATTTGG TACTCACGTG AAGAGGGAAA GCGGAAGATG GTAATTTTTT 240
TTTACGGATG ATATGGCAGG ATGATTGGTT CTGATCTTAC CGGCTAGTGG TCATTTTTAA 300
AAACTTGTA CCCTCTTATC TGAAATCTG TTTCTGGGAA TTGGCCATT TTAAGTGATT 360
TTGTTGCCC TCTTCTATNA ATATTCCTAC TTCCNTAAT AATGACTGAT TTNATTGTA 420
ANTCAGGNAT TTATNAAACC CTGGGGCTAC CAAGNCTTGT TAAACAT 467

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 442 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

:DOCID: WO 9844112A1

WO 98/44112

PCT/US98/06022

-82-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TACTGTTGGG GTGGCAGGAA ATTTCCAAAA TTGAAAACT GAATTCCACA ACTGACAAC	60
GAAATTCAG TTTTCTAGAC AATTTAAAT TTACCCAAAT CCCATAAAAA TAAACATAAT	120
TCCATTCAC GAAAACCACT GGATAATGTT TAACAAAGTC TTTGTAGCCA ATAGGTTTAT	180
AAATACCTGA ATTACAAATA AATCAGTCAT TATTATGGGA AGTAGGAATA TTATAGAAGA	240
GGGCAACAA AATCACTTAA AATGGCCAAA TTCCAGAAAC AGGATTCAG ATAAGAGGGT	300
ACAAGTTTT TAAAAATGAC CACTAGCCCG GTAAGATCAG AACCAATCAT CCTGCCATAT	360
CATCCGTAAA AAAAAATTAC CATCTCCGC TTTCCCTCTT CACGTGAGTA CCAAATTGGA	420
AGGGGTTAGA GGCCCNCCAA CG	442

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 575 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TACTGTTGGG GTCGGCAGGA AATTTCCACA ATTGAAAAAC TGAATTCCAC AACTGACAAC	60
TGAAATTTCA GTTTTCTAGA CAATTTAAAA TTTACCCAAA TCCATAAAAA ATAACATAA	120
ITCCATTTCA CGAAAACCAC TGGATAATGT TTAACAAAGT CTTGTAGCC AATAGGTTTA	180
TAAATACCTG AATTACAAAT AAATCAGTCA TTATTATGGG AAGTAGGAAT ATTATACGAA	240
GAGGGCAAAC AAAATCACTT AAAATGGCCA AATTCAGAA ACAGGATTC AGATAAGAGG	300
GTACAAGTTT TTTAAAAATG ACCACTAGCC CGGTAAGATC AGAACCAATC ATCCCTGGCC	360
ATATCATCCG GTAAAAAAA ATTACCATCT TCCGCTTTTC CCTCTTCAG TGAGGTACCC	420
AATTGGAANG GGTTTAGAAG GCCCTCAAAC GGGTATCACT CNTTAAAGGC ANTTTCATGG	480
GTTAATATGG AATTACCNCT AATGGTGAGA CCCACCTGG GGACTATTCC AAATACCCCT	540
TCCATGGCAA ATTGGNGTTG GAACCANTTT AGCAT	575

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 511 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both

WO 98/44112

PCT/US98/06022

-83-

(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
GGATACCCAT TTGAGGGCCT CTAAACCCCT CAATTTGGTA CTCACGTGAA GAGGGAAAGC      60
GGAAGATGGT AATTTTTTTT TATGGATGAT ATGGCAGGAT GATTGGTTCT GATCTTACCG      120
GCTAGTGGTC ATTTTAAAAA AACTTGTACC CTCATTCTG AAATCCTGTT TCTGGAATTT      180
GGCCATTTTA AGTGATTTTG TTTGCCCTCT TCTATAATAT TCCTACTTCC CATAATAATG      240
ACTGATTTAT TTGTAATTCA GGTATTTATA AACCTATTGG CTACAAAGAC TTTGTTAAAC      300
ATTATCCAGT GGTITTCGTG AAATGGGAAT TATGTTTATT TTTATGGGGA TTTGGGTAAA      360
TTTTAAATTG TCTAGGAAAA CTGAAATTTT CAGTTGTCCA GTTGTGGGAA TTCAGTTTTT      420
CCAATTGTGG GAAATTTCCT GGCCACCCCA ACAGTATTTT TGTGTGGTTA ATTAATTTTT      480
GCCAAATGAG GATCCNGGGT GTGACCACTN T                                511
```

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 400 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
GTAAATTTTA AATTGTCTAG AAAACTGAAA TTTCAGTTGT CAGTTGTGGA ATTCAATTTT      60
TCAATTGTGG AAATTTCTG CCACCCCAAC AGTATTTTTG TGTGTTAATT AATTTTGCAA      120
AATGAGAATC ATGGTGTGAC ACTCATCTAA TTTATCTTGT TGTGATGTTA TGGTCATAAT      180
AAGGAGAAAG AGGGTTTAAT TTTTCTTGTA TTTGGTTTCC TGGTGGTATC ATAGTGTAAT      240
TTTAGTATTT GAAATCAGT GTGATTCCTT AATGGGCCAA CTGAAGATTG AATTGCCGCT      300
AACCAACCATA TCGTGTTAGT GAATTTTCAA TATGGGACCN GGAAGGGCAT ATGTATTTTG      360
GAACTTGGAG TGGAAAAGGT TGGAGTTACA GACTTTTGGC                                400
```

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 480 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both

WO 98/44112

PCT/US98/06022

-84-

(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TGAGGGTCAC GTCGGTGGTT ACGGGTTTTA ACAACTTGCC GGACAGATTT AAAGACTTTC	60
TGCTGTATTT GAGATGCCGC AATTATTCAC TGCTTATAGA TCAGCCGGAT AAGTGTGCAA	120
AGAAACCTTT CTTGTTGCTG GCGATTAAGT CCCTCACTCC ACATTTTGCC AGAAGGCAAG	180
GCAATCCGGG AATCCTGGGG CCAAGAAAGC AACGCAGGGA ACCAAACGGT GGTGCGAGTC	240
TTCTGCTGG GCCAGACACC CCCAGAGGAC AACCACCCCG ACCTTTCAGA TATGCTGAAA	300
TTTGTAGAGT GAGAAGCACC AAGACATTCT TATGTGGGAA CTACAGAGGA CACTTTCCTT	360
CAANTTGTCT NTGGAAGGAA GTGCTGTTTT TTCAGGTGGG GTTAAGTTAT TTCCTGCCCC	420
GACATTGAGT TTGTTTTTTC AAGGGGCGAT GGACGATGTT TTTGTTGNAC ACCCTTCACT	480

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 392 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTAATTAACA CACAAAATA CTGTTGGGTA NANNAANAAA TTTCCACAAT TGAAAACTG	60
AATCCACAA CTGACAACTG AAATTTCACT TTTCTAGACA ATTTAAAATT TACCCAAATC	120
CCATAAAAT AACATAATT CCATTTCACTG AAAACCACTG GATAATGTTT AACAAAGTCT	180
TTGTAGCCAA TAGGTTTATA AATACCTGAN TTACAAATAA ATCAGTCATT ATTATGGGAA	240
GTAGGAATAT TATAGAAGAG GGCAAACAAA NTCACCTAAA ATGGCCAAAT TCCAGGANAC	300
AGGGATTTC AATAAGAGGG TACAAGTNTT TTAAAAGTGA CCACTAGGCC GGGTAAGGTC	360
CGGANCCAAT CATCTGCCA TTTTCATCCG TA	392

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

WO 98/44112

PCT/US98/06022

-85-

(11) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GNCCCCAGTA TCCCATCTGA TAAGAACCTT CAATTCTATA AACAAAAATA TTTCAAGAAA	60
GTATGTTACA CAATAGTACA TATAAGTAAT AGTTTGGCAG AATTTTAAAC TCTAGTAGTT	120
CATACCCCCA AAAAACAAAT TTTAAATTC AAAAATAACA GTTTTATTTA ACATATGTTA	180
CACCTTAACA TTTAAATAT CATGCTCTAG TTAAATATTT CATCAACAAC ACTGTATACA	240
ANTAAAATAT TACATAANAT ATATTTAAGG NAAATGTTTT GGGTCTTTGA TCTGGAACAN	300
TAAATAAAAA CACGGGCACT TCTACATAGG ACGGGGGTGG CGGTTACTAC TCCAATAATA	360
ATCNTGGTNT AGGGCGGCCT G	381

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 323 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GATCAAAGAC CAAACATTT TCTTAAATAT ATTTTATGTA ATATTTTATT TGTATACAGT	60
GTGTTGATG AAATATTTAA CTAGAGCATG ATATTTTAAA TGTTAAGGTG TAACATATGT	120
TAAATAAAAC TGTATTTTN GAATTNAAA ATTNNGTTTT NGGGGGTATG ANCTACTAGA	180
GTTTAAATTT CTGCCAACT ATTACTTATA TGTNCTATTG TGTAACATAC TTNCTNGAAA	240
TATTTNGGTT TATAGAATTG ANGGTTCTTA TCAGATGGGA TACTGGGGAC TATAACAAT	300
GGAAATAAAG CCACTGTATT TNT	323

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 299 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

WO 98/44112

PCT/US98/06012

-86-

AAAAATACAG TGGCTTTATT TCCATTGTTT ATAGTCCCCA GTATCCCATC TGATAAGAAC 60
CTTCAATTCT ATAAACAAAA ATATTTCAAG AAAGTATGTT ACACAATAGT ACATATAAGT 120
AATAGTTTGG CAGAATTTTA AACTCTAGTA GTTCATACCC CCAAAAAACA AATTTTAAAN 180
TTCAAAAATA ACAGTTTTAT TTAACATATG TTACACCTTA ACATTAAAA TATCATGCTC 240
TAGTTAATA TTTCATCAAC AACACTGTAT ACANNTAAAA TATTACATAA AATATATTT 299

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GACAGATTTA AAGACTTTCT GCTGTATTTG AGATGCCGCA ATTATTCAC TCTTATAGAT 60
CAGCCGGATA AGTGTGCAAA GAAACCTTTC TTGTTGCTGG CGATTAAGTC CCTCACTCCA 120
CATTTTGCCA GAAGGCAAGC AATCCGGGAA TCCTGGGGCC AAGAAAGCAA CGCAGGGAAC 180
CAAACGGTGG TCGAGTCTT CCTGCTGGGC CAGACACCCC CAGAGGACAA CCACCCCGAC 240
CTTTCAGATA TGCTGAAATT TGAGAGTTAG AAGCACCAAG ACATTCCTTA TGTGGGACCT 300
ACA 303

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 317 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TTTATTTCCA TTGTTTATAG TCCCCAGTAT CCCATCTGAT AAGAACCCTC AATTCTATAA 60
ACAAAAATAT TTCAAGAAAG TATGTTACAC AATAGTACAT ATAAGTAATA GTTTGGCAGA 120
ATTTTAAACT CTAGTAGTTC ATACCCCCAA AAAACAAATT TTAATTTCA AAAATAACAG 180
TTTTATTTAA CATATGTTAC ACCTTAACAT TTAAATATC ATGCTCTAGT TAAATATTTT 240
ATCAACAACA CTGTATACAA ATAAATATT ACATAAANTA TATTTAAGGN AATGTTTTG 300

WO 98/44112

PCT/US98/06022

-87-

GGTCTTTGAT CTGGAAC

317

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 325 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

TTAAAAATAN AGTGGCTTTA TTCCATTGT TTATAGTCCC CAGTATCCCA TCTGATAAGA	60
ACCTTCAATT CTATAACAA AAATATTTCA AGAAAGTATG TTACACAATA GTACATATAA	120
GTAATAGTTT GGCAGAATTT TAAACTCTAG TAGTTCATAC CCCCAGAAAA CAAATTTTAA	180
AATTCAAAAA TAACAGTTTT ATTTAACATA TGTTACACCT TAACATTAA AATATCATGC	240
TCTAGGTTAA ATATTTATC AACAACTG GTATACAAAT AAAATATTAC ATAAATATA	300
TTTAAGGGAA ATGTTTGGG GCTTT	325

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 282 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

TTTGGCCAGA AGGCAAGCAA TCCGGGAATC CTGGGGCCAA GAAAGCAACG CAGGGAACCA	60
AACGGTGGTG CGAGTNTTCC TGCTGGGCCA GACACCCCA GAGGACAACC ACCCCGACCT	120
TTCAGATATG CTGAAATTTG AGAGTGAGAA GCACCAAGAC ATTCTTATGT GGAACACAG	180
AGACACTTTN TTCAACTTGT CTCTGAAGGA AGTGCTGTTT CTNAGGTGGG TAAGTACTTC	240
CTGCCAGAC ACTGAGTTTG TTTTCAAGGG CGATGACGAT GT	282

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 358 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: both

WO 98/44112

PCT/US98/06022

-88-

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AATTTCCACA ATTGAAAAAC TGAATTCAC AACTGACAAC TGAAATTTC GTTTTCTAGA	60
CAATTTAAAA TTTACCCAAA TCCCATAAAA ATAAACATAA TTCCATTTC CGAAAACCAC	120
TGGATAATGT TTAACAAAGT CTTTGTAGCC AATAGGTTTA TAAATACCTG AATTACAAAT	180
AAATCAGTCA TTATTATGGG AAGTAGGAAT ATTATAGAAG AGGGCAAACA AAATCACTTA	240
AAATGGCCAA ATTCCAGGAA ACAGGGATT CAGGATAAGG GGGTACAAGT TTTTAAAAA	300
TGGACCACTA GGCCGGGTAA GGATCAGGAA CCANTTCATC CTGGCCATAT TCATCCGT	358

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 428 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ACTTCCTGCC CAGACACTGA GTTTGTTTTC AAGGGCGATG ACGATGTTTT TGTGAACACC	60
CATCACATCC TGAATTACTT GAATAGTTTA TCCAAGACCA AAGCCAAAGA TCTCTTCATA	120
GGTGATGTGA TCCACAATGC TGGACCTCAT CGGGATAAGA AGCTGAAGTA CTACATCCCA	180
GAAGTTGTTT ACTCTGGCCT CTACCCACCC TATGCAGGGG GAGGGGGGTT CCTCTACTCC	240
GGCCACCTGG GCCTGAGGCT GTACCATATT CACTGGACCA GGTCCATCT CTACCCCAT	300
TGGATGGACG TTTTATACTG GGAATGTGNC CTTCAGGAAA NTCGGGCCTC GTTTCAGGA	360
GGAAACACAA AGGTTTCAG GGGACATTTT GATATTCGAG GGGAGGGAAA AACAAAAANT	420
TAACATTT	428

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 266 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

150000:4WO_804112A1.JJ

WO 98/44112

PCT/US98/06022

-89-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TTTTGCCAGA AGGCAAGCAA TCCGGGAATC CTGGGGCCAA GAAAGCAACG CAGGGAACCA	60
AACGGTGGTG CGAGTNTTCC TGCTGGGCCA GACACCCCA GAGGACAACC ACCCCGACCT	120
TTCAGATATG CTGAAATTG AGAGTNAGAA GCACCAAGAC ATTCTTATGT GGAACACAG	180
AGACACTTTC TTCAACTTGT CTCTGAAGGA AGTGCTGTTT CTCAGGTGGG TAAGTACTTC	240
CTGCCCAGAC ACTGAGTTTG TTTTCA	266

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 259 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

TTGTTGGGTA TCCTGATGAT GGCAAATGTC TTCATTATT TTATTATGGA AGTCTCCAAA	60
AGCAGTAGCC AAGAAAAAAA TGGAAAAGGG GAAGTAATAA TACCCAAAGA GAAGTTCTGG	120
AAGATATCTA CCCCTCCCGA GGCATACTNG AACCGAGAGC AAGAGAAGCT GAACCGGCAG	180
TACAACCCCA TCCTGAGCAT GCTGACCAAC CAGACGGGGG AGGCGGGCAG GCTCTCCAAT	240
ATAAGNCATC TGAACACT	259

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 247 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TTATAGNCCC CAGTATCCCA TCTGATAAGA ACCTTCAATT CTATAACAA AAATATTTCA.	60
AGAAAGTATG TTACACAATA GTACATATAA GNAATAGTTT GGCAGAATTT TAACTCTAG	120
TAGTTCATAC CCCCACAAAA CAAATTTTAA AATTCAAAA TAACAGTTTT ATTTAACATA	180
TGTTACACCT TAACATTTAA AATATCATGC TCTNGTTAAA TATTCATCA ACAACACTGT	240
ATACAAA	247

WO 98/44112

PCT/US98/06022

-90-

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 368 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
CTGATGTTAG TACATAGTAG AAAACCTCAA GAGATGATTG ATATTGGTC TCAGTTGCAG      60
AGTGCTCATT TAAATGCTA AAATAGATAC AACTCAATT TTGCATAGAA AGGTGTATTT      120
TGAATAGTTC CCATGTTGTG TTCTCACATT AGAGTAATTT CTGTATTAAA CCATGAAAT      180
TGCCTTTATG AGTGATACCC ATTTGAGGGG CCTCTTAAAC CCTTCAATTT GGGTACTTCA      240
CGTGAAGAGG GGAAAGCGGG AAGATGGGTA ATTTTTTTTT ACGGGATGGA TATGGGCNGG      300
GATGATTGGG TTCTGGATCC TTACCCGGCC TAGTGGGTCC ATTTTTTAAA AACTTGGA      360
CCCCCNCC                                     368
```

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 195 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
AAAAATACAG TGGCTTTATT TCCATTGTTT ATAGTCCCCA GTATCCCATC TGATAAGAAC      60
CTNCAATTCT ATAAACAAA ATATTCAAG AAAGTATGTT ACACAATAGT ACATATAAGT      120
AATAGTTTGG CAGAATTTA AACTCTAGTA GTTCATACCC CCAAAAACA AATTTTAAA      180
TTCAAAAATA ACAGT                                     195
```

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2257 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: both
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

WO 98/44112

PCT/US98/06022

-91-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GAAATATTTT TGTTTATAGA ATTGAAGGTT CTTATCAGAT GGGATACTGG GGACTATAAA	60
CAATGGAAAT AAAGCCACTG TATTTTTTAAT TTTTGTGTA ATGTGTAATC TATAATCCTT	120
TTGTTTCCCA TATTGAGAA CATTTTTCCC TGAAAGAGGC CAGTTTCCTC CCCAGAAACC	180
ATTACAGTAG TGTTGAACTA TCACTGTCTC TCAGTGGGTC ATCCATCTTT GCATTTAAAA	240
TCCCCAAAGT GCTTTCCCAT TTAAAGTCTT TAAAGAAAAG TGAGAATATT TATTTATGCT	300
TCCATTTTCA GTGAGTATAA ATAATTTAAT TAGGGAGTGG TGTGGCATTG TAAAGATTGT	360
GTTATCCTAA GCCATTCTA TTTTGGAGTT TGTAGCCACA AAGATGAAAT ATAGAATCAG	420
CCTTGACTAC TCAATTTCTT TTCATAGACC CATGTTGAGA AGACACTACT AACGTCCAGT	480
GGGAAACAAG TAGACAATTG ATGAAGCTCA AAAAACAGAA GGGTTAGTGT TGTAAAGACA	540
AACAGTCTAA TCCTGTTTGG AATGTGGAAG CCATTTCTGA GCAAGTATGA GGACACAGGT	600
GCTTGATTG AGATTGAAGA CTGTTTTCAG CCTGGTCTTC CTGAAGGTTT CCTGGGGCCT	660
GCATCTGCC TCTACTCCCA TGGCTGCTAG CACACACCTC CCAGAGGGCC ATATTGCCAC	720
ATTATGGCTA GAGAAGAGTA AAGAAGAAAA GAAGCTCTGA GAACATTCAC AGGTAATTGG	780
ATCACATTTG CATTGTGCTA AAAAACCTGA CCACGCATTC TCAGGTAATA GGTTCCTCCT	840
CTCAGAGGAA TTTCAATTTT TTTTCTTGTT AGAGATTCCC CTTCTCTGAG GTTTCAAGTC	900
TCTTGTAGAG AAAGAAGAGA TGGAGCAGGT TTTGAATGAG GTGTGGAGGG CCACTGGGGG	960
GCCTTTTGTG AGCCTTCAGT CCACATGTGT GCTGTTGTTT GAACATGAGT TCTTGGTGCT	1020
GATGACATTT GGATGAGATG ATCTCTGGCC CTTCTTCATT TGGCAGAAGT TCTTGTGCAA	1080
TGGCTGCCCA AGCCCACCAC ACTGGTCATT GCTGCCCTGT GAGATGGACC TCATGGGCTT	1140
TTTAGCAGAG CACGTTAGGT TTTAGAGCTT TACGCATGCT TGGGCTCTGT TATGGCGCAA	1200
ACCCTTAAAT CCAGGAAGGC CTCTCTTGGT GCCACAATA TGGGTTCTCA CCTGATCCCC	1260
CATCTCACGG ATGGAACGTC TGTAAGTCTA ACTTATTCTT TGAGAACTGT TTAACAATTA	1320
GGCCTCAAGG GAAACTGGTA TTTTGGGCCC TTTTCTTGGC TATTCCCAAG TCATGTTGAT	1380
TTTGAGTTTG AAGGTCAAAA AGGCTGAAAG CATTGCCAGG GTTTGGACTA TTCAAAAACC	1440
CAAGCAGGTC TTAAAAAAG GATGCAAGAG ACAAGAATGG CTCATTCCCC TTCCTGATCC	1500
TGGTTATACC CATGTCCTTT CTTGAGATGG TCAAGAGAGG CTGGAAGAA GAACAGGAAA	1560
TTGGGGGAGT GCTTTGTAC ACTTGAAAT TGAGTCAAGA ATTAAAGACA CCCAAAGTGG	1620
GCCATCTCCT ACTTGTCCAC ACCTGATTGG TGGTGATGCG GAATATTTGA TGTCCCGGGT	1680
CATCTTGACT TTCTCAGATG CAAAAAGGGA GGGTGACTTT ACTAATGGAA AGGATGGGAA	1740

DOCID: WO_0844112A1_1

WO 98/44112

PCT/US98/06022

-92-

GCTGAAATGA ATGAAGCCTT CAGTTGGGCC AAAGTTTAAC TTCCCCGTGA TTGCCTTCT	1800
GATGAAAAAT GCCAGATGAA GTGAAAATTC TTGTTTCTTG CCTACAACAG GAAAATACAT	1860
ACTTTACATG CTGGGCTATT GAGGCTATGA AATTAGGTTT TCCTTAATGT AAATCCAATT	1920
GCTAGAAACA TTTGCCAAAT AAGATTTTTT GAACTGAACT TTGTTTGCAT TAATCTGAAA	1980
AACTGAAGTA TTCTGACTCA TGAAGTTCTC AAAGTAATAC ACTAAAAAG TTTTGCCCTT	2040
AATACCATTA TATCTTGTAG AGGCCAAGAA TGAGGGACTT CTGTCTTTAA AGAGCCCTAA	2100
AAATCTCGTT TGCTCACATG ATATGAATTA CCGTATTTGT TGTAAATGCG CAACTTTGTA	2160
TACACTAAAA GCAC TGCCAA TATGATTTTT TATCAGTTGT GCCTCAGTTA GAGATATTAA	2220
AATGTGACAT CTAAATATT ACATATTAGA ATAATTG	2257

WO 98/44112

-92.1-

PCT/US98/06022

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM
(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>3</u> , line <u>10</u> .	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit September 23, 1996	Accession Number ATCC 97729
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
DNA Plasmid, HTAAW41	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input type="checkbox"/> This sheet was received with the international application Specialist	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)

DOCID: <WO_9844112A1_1>

WO 98/44112

-92.2-

PCT/US98/06022

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM
(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>22</u> , line <u>22</u> .	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 25, 1998	Accession Number ATCC 209645
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
DNA Plasmid pHE4a	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g. "Accession Number of Deposit")	

For receiving Office use only	For International Bureau use only
<input type="checkbox"/> This sheet was received with the international application <div style="text-align: center; font-size: small;">RECEIVED PCT SECRETARIAT JAN 20 1999</div>	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer <div style="text-align: center; font-size: small;">RECEIVED PCT SECRETARIAT JAN 20 1999</div>	Authorized officer

Form PCT/RO/134 (July 1992)

NSDOCID: <WO_0844112A1_1>

WO 98/44112

PCT/US98/06022

93

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
 - 5 (a) a nucleotide sequence encoding a polypeptide comprising amino acids from about -32 to about 365 in SEQ ID NO:2;
 - (b) a nucleotide sequence encoding a polypeptide comprising amino acids from about -31 to about 365 in SEQ ID NO:2;
 - (c) a nucleotide sequence encoding a polypeptide comprising
10 amino acids from about 1 to about 365 in SEQ ID NO:2;
 - (d) a nucleotide sequence encoding a polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; and
 - 15 (e) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c) or (d).
2. An isolated nucleic acid molecule comprising a polynucleotide which encodes the amino acid sequence of an epitope-bearing portion of an CAPP polypeptide having an amino acid sequence in (a), (b), (c) or (d) of claim 1.
- 20 3. The isolated nucleic acid molecule of claim 1, which encodes an epitope-bearing portion of a CAPP polypeptide selected from the group consisting of: a polypeptide comprising amino acid residues from about -32 to about -22 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about -4 to about 40 in SEQ ID NO:2; a polypeptide comprising amino acid
25 residues from about 46 to about 57 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 62 to about 73 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 78 to about 87 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 92 to about 110 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 119 to about

3DOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

94

144 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 152 to about 186 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 200 to about 219 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 230 to about 240 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 248 to about 258 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 314 to about 336 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 344 to about 353 in SEQ ID NO:2.

10 4. An isolated nucleic acid molecule, comprising a polynucleotide having a sequence selected from the group consisting of:

(a) the nucleotide sequence of a fragment of the sequence shown in SEQ ID NO:1, wherein said fragment comprises at least 50 contiguous nucleotides of SEQ ID NO:1; and

15 (b) a nucleotide sequence complementary to a nucleotide sequence in (a).

5. A method for making a recombinant vector comprising inserting an isolated nucleic acid molecule of claim 1 into a vector.

20 6. A recombinant vector produced by the method of claim 5.

7. A method of making a recombinant host cell comprising introducing the recombinant vector of claim 6 into a host cell.

8. A recombinant host cell produced by the method of claim 7.

25 9. A recombinant method for producing any of the CAPP polypeptides, comprising culturing the recombinant host cell of claim 8 under conditions such that said polypeptide is expressed and recovering said polypeptide.

WO 98/44112

PCT/US98/06022

95

10. An isolated CAPP polypeptide having an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

- 5 (a) amino acids from about -32 to about 365 in SEQ ID NO:2;
(b) amino acids from about -31 to about 365 in SEQ ID NO:2;
(c) amino acids from about 1 to about 365 in SEQ ID NO:2;
(d) the amino acid sequence of the CAPP polypeptide having the amino acid sequence encoded by the cDNA clones contained in ATCC Deposit No. 97729; and
10 (e) the amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a), (b), (c) or (d).

11. An isolated polypeptide comprising an epitope-bearing portion of the CAPP protein, wherein said portion is selected from the group consisting of:
15 a polypeptide comprising amino acid residues from about -32 to about -22 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about -4 to about 40 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 46 to about 57 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 62 to about 73 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 78 to about 87 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 92 to about 110 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 119 to about 144 in SEQ ID NO:2;
20 a polypeptide comprising amino acid residues from about 152 to about 186 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 200 to about 219 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 230 to about 240 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 248 to about 258 in SEQ ID NO:2; a polypeptide comprising amino acid residues from about 314 to about 336 in SEQ ID NO:2; and a polypeptide comprising amino acid residues from about 344 to about 353 in SEQ ID NO:2.
25

IDOCID: <WO_9844112A1_1>

WO 98/44112

PCT/US98/06022

96

12. An isolated antibody that binds specifically to a CAPP polypeptide of claim 10.

13. An isolated nucleic acid molecule comprising a polynucleotide encoding a CAPP polypeptide wherein, except for one to fifty conservative amino acid substitutions, said polypeptide has a sequence selected from the group consisting of:

- (a) a nucleotide sequence encoding a polypeptide comprising amino acids from about -32 to about 365 in SEQ ID NO:2;
- (b) a nucleotide sequence encoding a polypeptide comprising amino acids from about -31 to about 365 in SEQ ID NO:2;
- (c) a nucleotide sequence encoding a polypeptide comprising amino acids from about 1 to about 365 in SEQ ID NO:2;
- (d) a nucleotide sequence encoding a polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; and
- (e) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c) or (d).

14. An isolated CAPP polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has a sequence selected from the group consisting of:

- (a) amino acids from about -32 to about 365 in SEQ ID NO:2;
- (b) amino acids from about -31 to about 365 in SEQ ID NO:2;
- (c) amino acids from about 1 to about 365 in SEQ ID NO:2;
- (d) the amino acid sequence of the CAPP polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97729; and
- (e) the amino acid sequence of an epitope-bearing portion of any one of the polypeptides of (a), (b), (c) or (d).

WO 98/44112

PCT/US98/06022

97

15. An isolated antibody that binds specifically to a CAPP polypeptide of claim 10.

DOCID: <WO_9844112A1_1>

WO 98/44112

1/7

PCT/US98/06022

Figure 1A

10 30 50
-231 GCAGCGGCACCGCAGCAGCGGCAACAAGTCCCAGACTAGCAGAGCCAAGCCGGAGCACTC -172

70 90 110
-171 CCTGCCGCGGACACCGCCGGGCGGCCCGTCCGGGGCGCCGCGCATGGAGCGTGGAGCTGCG -112

130 150 170
-111 GCGGTGCGCGGGGCTGAGCCGCGCGGAGCGCCGGGACGTGGATGTGGCCGCGATCTCCCG -52

190 210 230
-51 CCCTTGCCCCCGCCCCGAGCTGGAGCTGCTCCCGGACAAGATATGAGAAATGAGTGT 8
-16 M S V 3

250 270 290
9 TGGACGTGGAAGAATAAAGTTGTTGGGTATCCTGATGATGGCAAATGTCTTCATTTATTT 68
4 G R R R I K L L G I L M M A N V F I Y F 23

310 330 350
69 TATTATGGAAGTCTCCAAAGCAGTAGCCAGAAAAAATGGAAGGGGAAGTAATAAT 128
24 I M E V S K S S S Q E K N G K G E V I I 43

370 390 410
129 ACCCAAGAGAGAAGTTCTGGAAGATATCTACCCCTCCCGAGGCATACTGGAACCGAGAGCA 188
44 P K E K F W K I S T P P E A Y W N R E Q 63

430 450 470
189 AGAGAAGCTGAACCGGCAGTACAACCCCATCCTGAGCATGCTGACCAACCGAGACGGGGA 248
64 E K L N R Q Y N P I L S M L T N Q T G E 83

490 510 530
249 GCGCGGCAGGCTCTCCAATATAAGCCATCTGAAGTACTGCGAACCTGACCTGAGGGTCA 308
84 A G R L S N I S R L N Y C E P D L R V T 103

550 570 590
309 GTCGGTGGTTACGGGTTTACCAACTGCGCGGACAGATTTAAAGACTTTCTGCTGTATTT 368
104 S V V T G F N N L P D R F K D F L L Y L 123

610 630 650
369 GAGATGCGCAATTATTCACTGCTTATAGATCAGCCGATAAGTGTGCAAGAAACCTTT 428
124 R C R N Y S L L I D Q P D K C A K K P F 143

670 690 710
429 CTGTGTGCTGGCGATTAAAGTCCCTCACTCCACATTTTGCCAGAAGGCAGCAATCCGGGA 488
144 L L L A I K S L T P E F A R R Q A I R E 163

730 750 770
489 ATCCTGGGGCCAGAAAGCAACGCGGGAACCAACGGTGGTGGAGTCTTCTGCTGGG 548
164 S W G Q E S N A G N Q T V V R V F L L G 183

790 810 830
549 CCAGACACCCCGAGAGGACAACCCCGGACCTTTCAGATATGCTGAAATTTGAGAGTGA 608
184 Q T P P E D N H P D L S D M L K P E S E 203

SDOCID: <WO_9844112A1_1>

WO 98/44112

2/7

PCT/US98/06022

Figure 1B

609	850	870	890	668
204	GAAGCACCAAGACATTCTTATGTGGAACACAGAGACACTTTCTTCAACTTGTCTCTGAA			223
	K H Q D I L M W N Y R D T F F N L S L K			
669	910	930	950	728
224	GGAAGTGTCTGTTTCTCAGGTGGGTAAGTACTTCCTGCCCAGACACTGAGTTTGTCTTCAA			243
	E V L F L R W V S T S C P D T E F V F K			
729	970	990	1010	788
244	GGGCGATGACCATGTTTTTGTGAACACCCATCACATCCTGAATTACTTGAATAGTITATC			263
	G D D D V F V N T H H I L N Y L N S L S			
789	1030	1050	1070	848
264	CAAGACCAAGCCAAAGATCTCTTCATAGGTGATGTGATCCACAATGCTGGACCTCATCG			283
	K T K A K D L F I G D V I H N A G P H R			
849	1090	1110	1130	908
284	GGATAAGAAGCTGAAGTACTACATCCCAGAAGTTGTTTACTCTGGCCTCTACCCACCCTA			303
	D K K L K Y Y I P E V V Y S G L Y P F Y			
909	1150	1170	1190	968
304	TGCAGGGGGAGGGGGGTTCTCTACTCCGGCCACCTGGCCCTGAGGCTGTACCATATCAC			323
	A G G G G F L Y S G H L A L R L Y H I T			
969	1210	1230	1250	1028
324	TGACCAGGTCCATCTCTACCCCATTTGATGACCTTTTACTTGAATGTGCTTCAGAACT			343
	D Q V H L Y P I D D V Y T G M C L Q K L			
1029	1270	1290	1310	1088
344	CGGCCTCGTTCCAGAGAAACACAAGGCTTCAGGACATTTGATATCGAGGAGAAAAACAA			363
	G L V P E K H K G F R T F D I E E K N K			
1089	1330	1350	1370	1148
364	AAATAACATCTGCTCCTATGTAGATCTGATGTTAGTACATAGTAGAAAACCTCAAGAGAT			383
	N N I C S Y V D L M L V H S R K P Q E M			
1149	1390	1410	1430	1208
384	GATTGATATTTGGTCTCAGTTGCAGAGTGCTCAITTTAAAATGCTAAAATAGATACAACT			397
	I D I W S Q L Q S A H L K C *			
1209	1450	1470	1490	1268
	CAATTTTGCATAGAAAGGTGTATTTTGAATAGTTCCCATGTTGTGTTCTCACATTAGAGT			
1269	1510	1530	1550	1328
	AATTTCTATATTAAACCATGAAAATTGGCTTTATGAGTGATACCCATTGAGGGGCTCTA			
1329	1570	1590	1610	1388
	AACCCCTCAATTTGGTACTCACGTGAAGAGGAAAGCGGAAGATGGTAATTTTTTTTAT			
1389	1630	1650	1670	1448
	GGATGATATGGCAGGATGATTGGTTCTGATCTTACCGGCTAGTGGTCAITTTTAAAAAAC			
1449	1690	1710	1730	1508
	TTGTACCCTCTTATCTGAAATCCTGTTTCTGGAATTTGGCCATTTTAAAGTGATTTTGTIT			

DOCID: 4WO_9844112A1_2

WO 98/44112

3/7

PCT/US98/06022

Figure 1C

1509 1750 1770 1790 1568
GCCCTCTTCTATAATATTCCTACTTCCCATATAATGACTGATTTATTTGTAAATTCAGGT

1569 1810 1830 1850 1628
ATTTATAAACCTATTGGCTACAAAGACTTTGTTAAACATTATCCAGTGGTTTTCGTGAAA

1629 1870 1890 1910 1688
TGGAAATTATGTTATTTTATGGGATTGGGTAAATTTTAAATGTCTAGAAAACCTGAAA

1689 1930 1950 1970 1748
TTTCAGTTGTTCAGTTGTGGAATTCAGTTTTTCAATTGTGGAAATTCCTGCCACCCCAAC

1749 1990 2010 2030 1808
AGTATTTTGTGTGTTAATTAATTTTGCAAAATGAGAATCATGGTGTGACACTCATCTAA

1809 2050 2070 2090 1868
TTTATCTTGTGTGATGTTATGGTCATAATAAGGAGAAAGAGGGTTAATTTTCTTGTA

1869 2110 2130 2150 1928
TTTGGTTTCTGGTGGTATCATAGTGTAAATTTAGTATTTGAAAATCAGTGTGATTCCTT

1929 2170 2190 2210 1988
AATGGCCAACTGAAGATTGAATTGCGGCTAACAACCATATCGTGTAGTGAATTTTCAAT

1989 2230 2250 2270 2048
ATGGACCAGGAAGGCATATGTATTTTGAACCTGAGTGAAAAGGTTGAAGTTACAGACTTT

2049 2290 2310 2330 2108
TGCATAGATGGTTTGTCAATTTAAATTTCCAGAAATTTATTATGCCATATTTTCACATGC

2109 2350 2370 2390 2168
TGCTTATACAGATTATTATTGAGTAGTAAGTGTCCCTGTCTATGTAGAAGTGCCCTGTG

2169 2410 2430 2450 2228
TTTTTATTTATGTTCCAGATCAAAGACCRAAACATTTCTTAAATATCTCTTATGTAAT

2229 2470 2490 2510 2288
ATTTTATTTGTATACAGTGTGTTGATGAAATATTTAACTAGAGCATGATATTTTAAATG

2289 2530 2550 2570 2348
TTAAGGTGTAAACATATGTTAAATAAACTGTTATTTTGAATTTTAAATTTGTTTTTG

2349 2590 2610 2630 2408
GGGGTATGAACTACTAGAGTTTAAATTTCTGCCAACTATTTACTTATATGTACTATTGTG

2409 2650 2670 2690 2468
TAACATACTTTCTTGAAATATTTTGTGTTATAGAATTGAAGGTTCTTATCAGATGGGATA

2469 2710 2730 2513
CTGGGGATTATAAACAAATGGAATAAAGCCACTGTATTTTAAAA

WO 98/44112

PCT/US98/06022

4/7

FIG. 2

```

60 NREQEKLNRQYNPILSMLTNQTGEAGRLSNISHLNYCEPDLRVTSVVTGF 109
   . . . . | | . : : | . : . : . | | . : . | | : | . : . . .
2  QSKHRKLLLRCLLVLPIL.LVDYCGLLTHLHELNF.ERHFHYPLNDDTG 49

110 NNLPDRFKDFLLYLRCRNYS..LLIDQPDKCAKPFLLLAIKSLTPHFAR 157
   . . . . | : | | . . . . : : | | . . . . | : | | . . . . |
50 SGSASSGLDKFAYLRVPSFTAEPVDPAR.....LTMLIKSAVGNSRR 93

158 RQAIRESWGQESNAGNQTIVVRVFLGQTPPEDNHPDLSDMLKPFSEKHQD 207
   | : | | . | | : . : : | | | | . . . . . : : | | . | . |
94 REAIRRTWGYEGRPSDVHLRRVFLGTAEDSEKD.....VAVESREHGD 137

208 ILMWNYRDTFFNLSLKEVLEPLRWVSTSCPDTEFVFKGDDDDVFVNTHHILN 257
   | | : : | . : | | . | | : | : | | . . . . . | | : . | | : | . : : | .
138 ILQADFTDAYFNNTILKTM LGMRWASEQFNRSEFYLFVDDDDYYVSAKNVLK 187

258 YLNSLSKTKAKDLFIGDVIHNAGPHRDKKLKYIPEVVYS.GLYPPYAGG 306
   : | . . . . . : | : : : : . : | | . | | : | : . | : : : | | . :
188 FLGRGRQSHQPELLFAGHVFTSPLRHKPSKQWYVSLEEYPPDRWPPYVTA 237

307 GGFLYSGHLALRLYHITDOVHLYPIDDVYTGMCLOKLGLVPEKHKGFRTF 356
   | : | : . | . . | | . : : | : : | | | | | | . : . | | : : . : | |
238 GAFILSOKALRQLYAASVHLPLFRFDDVYLGIVAALKAGISLQHCDDFRFH 287

357 DIEEKNKNNICSYVDLMLVHSRKPOEMIDIWSQLOSAHL 395
   . | . . . . : | : . . . : | | . | : . : | | : .
288 RPAYKGPDSSYSSVIA..SHEFGDPEEMTRVWNECRSANY 324

```

Symbol comparison table: /gcg/gcgcore/data/rundata/swgappep.cmp
CompCheck: 1254

Gap Weight:	4.000	Average Match:	0.540
Length Weight:	0.200	Average Mismatch:	-0.396

Quality:	162.9	Length:	339
Ratio:	0.504	Gaps:	7
Percent Similarity:	48.438	Percent Identity:	27.812

nsp100.pep x brn.pep May 22, 1996 11:08 ..

WO 98/44112

PCT/US98/06022

5/7

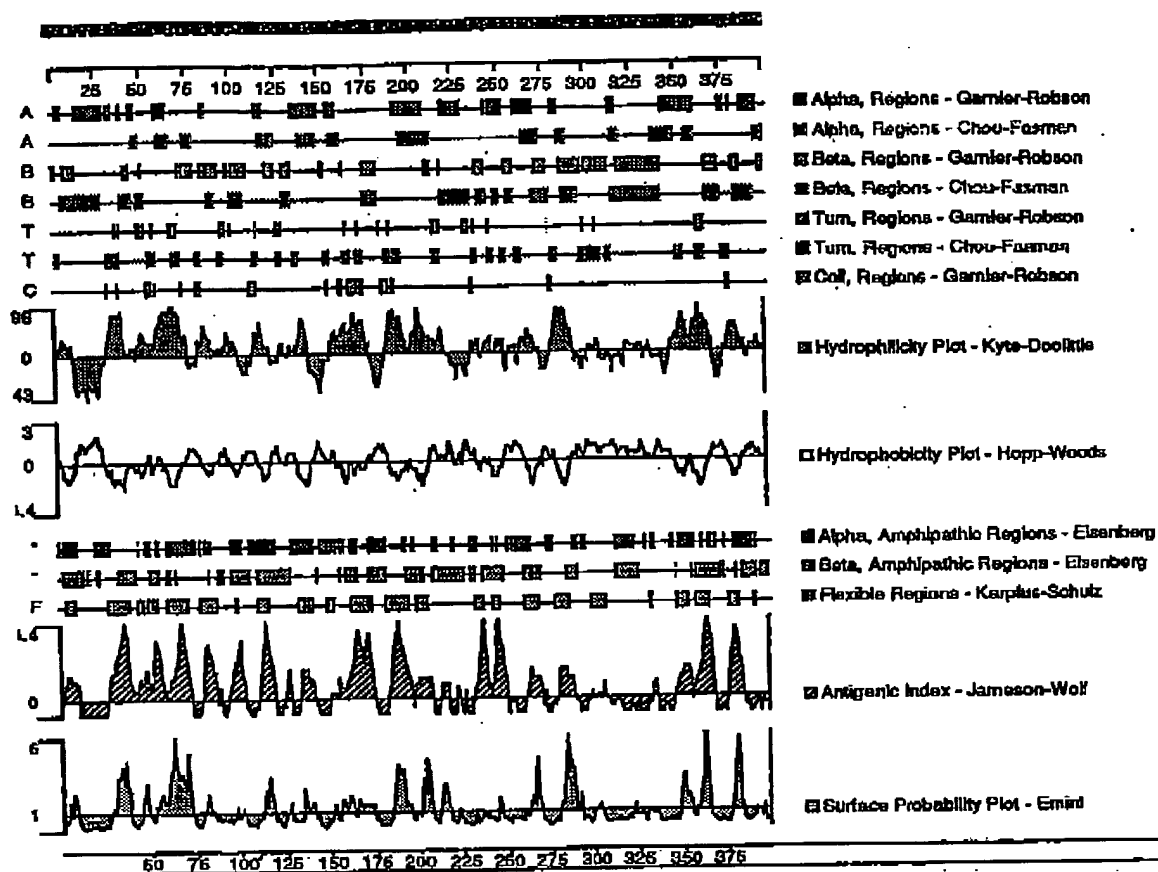


Figure 3

WO 98/44112

PCT/US98/06022

6/7

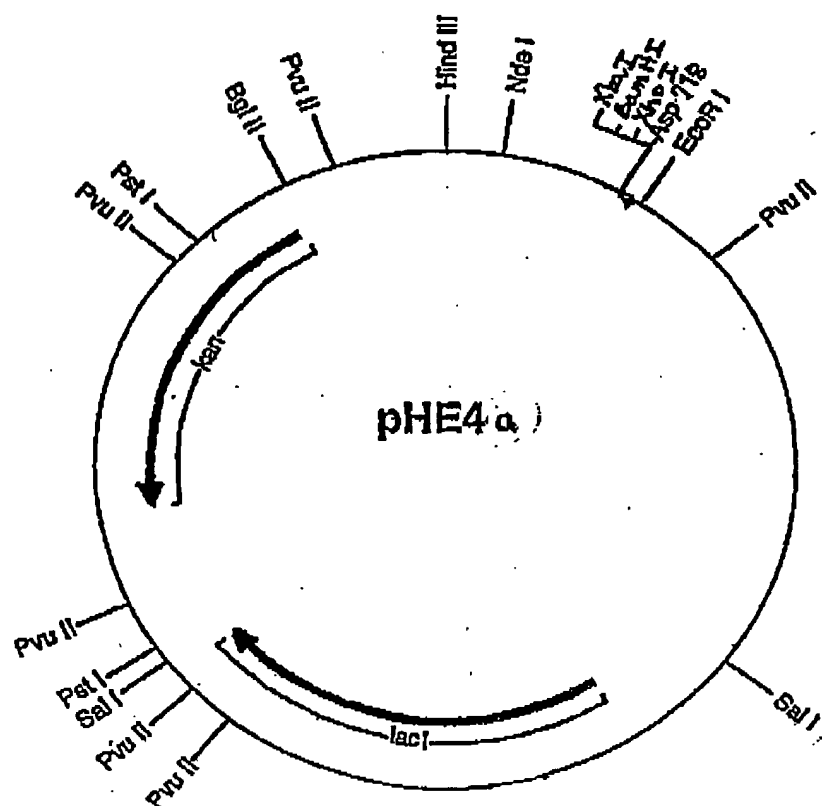


FIG. 4

DOCID: <WO_8844112A1_>

